Uncontrollable: Data subject rights and the data-driven economy

Ursic, H.

Author: Ursic, H.
Title: Uncontrollable: Data subject rights and the data-driven economy
Issue Date: 2019-02-07

2. THE RISE OF THE DATA-DRIVEN ECONOMY AND THE INDIVIDUAL

2.1. Introduction

Some see the 21st century as a new industrial and economic era.[60] In their view, the recent technological developments – social media, connected devices, datafication, and ubiquitous computing – have been so significant that they have opened the way for the next technological revolution and set the basis of an entirely new industry.[61] As Klaus Schwab, the founder of the World Economic Forum, puts it, *'a fusion of technologies that is blurring the lines between the physical, digital and biological spheres has characterized a fourth industrial revolution …'* .[62]

The change between the world in the 1990s and the world of today is undoubtedly apparent. Consider a simple example: a TV device. In the 1990s, almost every household owned a large, clumsy, and heavy device which emitted a high level of electromagnetic radiation and offered, by today's standards, the bare minimum of picture quality. Today, users have witnessed a new generation of televisions: devices with flat screens. Moreover, a TV is no longer solely a TV. Shows, movies, video games, apps, streaming, and more – all of this is available to a TV user simply by swiping on a touch screen. TVs have become smart devices that understand users' wishes and communicate with other devices. However, TVs are also smart for another reason: they are able to work behind the scenes. Ceaselessly and quietly, TVs collaborate with their manufacturers and share data about users' watching habits. The same data can later be sold to a third party, e.g. an advertiser. Vizio, a California-based TV maker, was recently found to follow such commercial tactics.[63] The company tracked customers in a way that enabled it to connect their viewing habits to their IP addresses.[64] Advertisers that were given access to this data were able to target users through several different mobile devices.[65]

Regardless of whether data-driven strategies are seen as a new stage in the history of the world's industry or as a continuation of the digital revolution, one thing is certain: the role of data in the economy is becoming much more pervasive. As demonstrated by the example above, the possibility of

---

[60] See for example Dirk Helbing, 'Economy 4.0 and Digital Society: The Participatory Market Society Is Born (Chapter 8 of Digital Society)' [2014] <http://papers.ssrn.com.ezproxy.liv.ac.uk/sol3/papers.cfm?abstract_id=2539330> accessed 27 May 2018.

[61] Klaus Schwab, 'The Fourth Industrial Revolution: what it means, how to respond' World Economic Forum (14 January 2016) <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/> accessed on 27 May 2018.

[62] Ibid.

[63] Ellie Zolfagharifard, 'Is YOUR TV spying on you? Report reveals how Vizio smart televisions track your data so that it can be sold to advertisers', *DailyMail* (10 November 2015) <http://www.dailymail.co.uk/sciencetech/article-3312597/Is-TV-spying-Report-reveals-Vizio-smart-televisions-track-data-sold-advertisers.html#ixzz4Ij0DiJQe> accessed on October 15, 2016.

[64] '*An Internet Protocol address ("IP addres") is a sequence of binary numbers which, when allocated to a device (a computer, a tablet or a smartphone), identifies it and allows it to access that electronic communications network. The device, in order to connect to the Internet, must use the number sequence provided by Internet service providers. The IP address is transmitted to the server on which the accessed web page is stored*.' C-582/14 *Patrick Breyer v Bundes republik Deutschland* [2016] ECLI:EU:C:2016:339, Opinion of AG Campos Sánchez-Bordona, para. 1.

[65] In 2013 LG advertised its smart TV by praising its ability to target advertising based on the profiles built on the data that is collected during the use of the TV: '*LG Smart Ad can feature sharp suits to men, or alluring cosmetics and fragrances to women.'* 'LG Smart TVs logging USB filenames and viewing info to LG servers' (*Blogspotcom*, 18 November 2013) <http://doctorbeet.blogspot.com/2013/11/lg-smart-tvs-logging-usb-filenames-and.html> accessed 27 May 2018.

secondary uses gives big data a new function and opens up business opportunities that were previously unimaginable.

Despite its great potential, big data use is not always innocent and beneficial to individuals. From the latter's perspective, its use can also be risky, ethically disputable, and sometimes illegal. The following sections of this chapter provide examples of all three situations.

By exploring the drivers of big data and the position of individuals in the data-driven economy, Chapter 2 answers the first sub-question of this thesis: *What are the driving forces behind the rise of the data-driven economy, taking into consideration technological enablers, economic consequences, and their interdependencies, and what are the consequences for individuals?* For the sake of clarity, the answer is split into three sections.

Section 2.2. focuses on three technological enablers that have led to the data revolution: the Internet, datafication, improved storage capabilities, and analytics. The aim here is not to go into technical details but to provide some background information that will be useful in the further analysis of legal responses to technological developments. As the intention is to illustrate the role individuals play in the data economy, technical developments are only explained to the degree that allows a meaningful demonstration of possible impacts on individuals. Indeed, even more than written norms imposed by a regulator, technology can strengthen or weaken an individual's position, which is a plausible reason for a brief analysis.[66]

After the technology-focused section, section 2.3. explains how the data-driven economy works. To simplify the complex economic ecosystem, an illustration of a data value chain consisting of three links is used: 1) data generation and acquisition, 2) analysis of data, and 3) data-driven decision-making. The key point is to understand why, from a business perspective, (personal) data can be a useful source of information and how the use of (personal) data has influenced the economy.

In the final section (2.4.), the focus is on individuals as an important group of actors in the data economy. It is shown how the data economy can work to both their advantage and disadvantage. Knowing more about risks and benefits is critical before taking on a legal discussion, which shifts the focus to addressing and mitigating the risks.

2.2. Technologies that created the data-driven economy

The sections below reflect upon four technological pillars that form the foundation of the data economy: the Internet, indefinite data storage, datafication, and data analytics. Undoubtedly, many more technologies have contributed to the rapid change in the data economy, such as advanced software and networks infrastructure. However, as the scope of this chapter is limited, these are not explained in more detail.

[66] Lawrence Lessig, *Code: Version 2.0* (Basic Books 2006). For example, in a recent interview Lessig praises Enigma, a decentralized cloud platform based on blockchain technology, for its privacy-enhancing invention. Their system could address problems which laws cannot. Steve Rosenbush, 'Lawrence Lessig: Technology Will Create New Models for Privacy Regulation' *The Wall Street Journal* (30 December 2015), <https://www.wsj.com/news/cio-journal> accessed 27 May 2018.

2.2.1. Internet (of Things)

The Internet or, more specifically, the World Wide Web is one of the fundaments of the data economy.[67] Its foundations were laid by Tim Berners-Lee's invention of the hypertext markup language in 1989, which created an open system platform where data and information could be shared and accessed instantaneously across the world.[68] This opened up new possibilities for the economy.[69] Today, the power of the Internet is amplified due to the rapid diffusion of broadband creating the underlying infrastructure for the exchange and free flow of data.[70] Data collected remotely through Internet applications and now increasingly through smart and interconnected devices can be easily shared and transferred all around the world.[71]

Not only the industry has benefited from the Internet: users who were once passive have also been given exciting opportunities to take a more active role on the Internet.[72] By using modern and greatly improved technologies, they have become more directly involved in content generation and sharing.[73] This user-friendly and social networking web, also known as Web 2.0,[74] has led to an increased opportunity to capture personal data. The more active the users are, the richer the data they leave behind.

The Internet is still evolving. The latest era in the Internet history is the so-called Internet of Things (IoT) or ubiquitous computing.[75] IoT stands for 'things' such as devices or sensors that connect, communicate, or transmit information with or between each other through the Internet.[76] IoT is fuelled by the prevalence of devices enabled by open wireless technology such as Bluetooth, radio-frequency identification,[77] Wi-Fi,[78] and telephonic data services, along with embedded sensors.[79] These machines

---

[67] Many people use the words 'Internet' and 'the World Wide Web' interchangeably. While they are indeed linked, they are two separate phenomena. The Internet is a global system of interconnected computer networks using the Internet protocol suite (TCP/IP) to link devices. The World Wide Web ('www' or 'web' for short) is a set of protocols and conventions which creates a universe of network-accessible information. The web is easy for anyone to roam, browse, and contribute to through the use hypertext and multimedia techniques, and can be accessed via the Internet by using web browsers such as Google Chrome, Internet Explorer or Mozilla Firefox. Lessig (2006) 145-146. Also see 'About World Wide Web' W3C, 24 January 2001 <https://www.w3.org/WWW/> accessed 14 June 2018.
[68] Lessig (2006) 146.
[69] Leo Bartevyan, 'Industry 4.0 – Summary Report' (2015) <https://www.cenit.com/fileadmin/dam/Corporate/PDFs/2015_5_Expertenwissen_E.pdf> accessed 27 May 2018.
[70] OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 35. Broadband is a high-speed communications network and especially one in which a frequency range is divided into multiple independent channels for simultaneous transmission of signals (as voice, data, or video). Definition from Merriam-Webster online dictionary <https://www.merriam-webster.com/dictionary/broadband> accessed on December 28, 2016.
[71] OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 35.
[72] Angela Daly, 'The Internet, User Autonomy and EU Law' (2016) <https://www.ssrn.com/abstract=2780789> accessed 27 May 2018.
[73] Bartevyan (2015).
[74] Viktor Mayer-Schönberger, *Delete: The Virtue of Forgetting in the Digital Age* (Princeton: Princeton University Press, 2009).
[75] Sometimes also called the Internet 3.0 or 4.0., see for example J Gubbi and others, 'Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions' (2013) 29 Future Generation Computer Systems 1645.
[76] FTC, 'Internet of Things: Privacy & Security in a Connected World' (2015) 5.
[77] A tag containing a unique ID. If it is destined to be carried by a person, then the tag ID should be considered as personal data. Article 29 Data Protection Working Party, '*Opinion 9/2011 on the Revised Industry Proposal for a Privacy and Data Protection Impact Assessment Framework for RFID Applications*' (2011).
[78] In particular the wireless sensor network, which stands for a network of nodes that cooperatively sense and control the environment, enabling interactions between persons or computers and the surrounding environment.
[79] Gubbi and others (2013).

have the potential to produce more and more data and to create new opportunities for their employment and application. Personal data represents a huge amount of the data captured in the IoT.

IoT indicates an interesting shift in the history of the Internet. While Web 1.0 was not yet developed to the degree that would enable *individuals* to be active users, Web 2.0 changed this as it gave individuals tools to be more engaged online. In the era of IoT or Web 3.0, the trend has reversed. Individuals are becoming less involved because data processing typically takes place behind the scenes, e.g. by generating information through multiple omnipresent sensors. In recent years, the IoT has become so sophisticated that a random user hardly notices it. In fact, it is the point of IoT to be hidden from a user's view. What is also revolutionary is that these physical information systems are now starting to be deployed, and some of them even work largely without human intervention.[80] Nevertheless, individuals can still be deeply involved in and dependent on these processes.[81] As mentioned above, data collected in the IoT environment very often relates, directly or indirectly, to individuals. For instance, after discussing babies in front of Amazon's personal assistant Alexa, a husband started receiving advertisements for Seventh Generation diapers on his Amazon Kindle.[82] Clearly, the devices were somehow connected, most likely via his email address. Having analysed the exchanged data, the device was able to predict the couple's highly personal plans and wishes.

2.2.2. Datafication

Over the past decade, digital data production and storage have grown exponentially.[83] However, another process has run in parallel to data increase: digitalised data has been translated into discrete, machine-readable, measurable, manipulable bits and bytes.[84] Putting data into a quantified format has enabled its tabulation and analysis. This transformation is commonly referred to as *datafication*.[85] The process of datafication started in the financial, energy, and retail industries,[86] but it soon expanded to other areas. In recent years, we have witnessed the datafication of words, markets, locations, and even human interactions.[87]

Individuals engage in the process of datafication by contributing their personal data as raw material. Every online appearance leaves a digital trace, which gradually grows into a vast registry of the actions constituting 'data doubles' or 'quantified selves'.[88] The process of *personal datafication* – a

---

80 McKinsey, 'Big Data: The next Frontier for Innovation, Competition, and Productivity'.
81 Rebecca Crootof, 'An Internet of Torts' (2018) <https://conferences.law.stanford.edu/werobot/wp-content/uploads/sites/47/2018/02/Crootof-An-Internet-of-Torts-We-Robot-Submission.pdf>.
82 Rory Carroll, 'Goodbye privacy, hello 'Alexa': Amazon Echo, the home robot who hears it all' *The Guardian* (21 November 2015) <https://www.theguardian.com/technology/2015/nov/21/amazon-echo-alexa-home-robot-privacy-cloud accessed> 27 May 2018.
83 Peter Géczy, 'Data Economy Dimensions' (2015) 9 Global Journal of Business Research.
84 Mireille Hildebrandt, 'Slaves to Big Data. Or Are We?' [2013] IDP Revista De Internet, Derecho I Política 6.
85 To illustrate the distinction between digitalisation and datafication we can use the example of a book copy. When the original book is scanned to make a digital copy, this is called digitization. Datafication of a book goes a step further by making the text indexable and thus searchable. Mayer-Schönberger and Cukier (2014) 144.
86 Jeff Bertolucci, 'Big Data's New Buzzword' *InformationWeek* (25 February 2013) <http://www.informationweek.com/big-data/big-data-analytics/big-datas-new-buzzword-datafication/d/d-id/1108797> accessed on 25 May 2018.
87 Shoshana Zuboff, 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization' (2015) 30 Journal of Information Technology 75, 79.
88 Gemma Galdon Clavell, 'Policing, Big Data and the Commodification of Security' in Bart Van Der Sloot, Dennis Broeders and Erik Schrijvers (eds), *Exploring the Boundaries of Big Data* (Amsterdam University Press 2016) 106. Also see Sara M. Watson, 'How Close Does Personalized Online Advertising Get to us as Our Real Persons?' *Schirnmag* (21 May 2016) <http://www.schirn.de/en/magazine/context/sara_m_watson_bits_of_me_essay/> accessed on October 3, 2016. Watson

transformation into data of multiple aspects of the lives of individuals[89] – has been the focus of a growing number of consumer-centred companies. Large Internet-based firms such as Google and Facebook are the most obvious example of the trend, but smaller Internet start-ups do not lag far behind.[90]

Datafication of personal information and behaviour is inherent to the process of personal data commodification and commercialisation. Commodification means exchanging data for something else, thus transforming datafied information into (monetary) value. Taking the form of a commodity, personal data is devaluated to the level of a commercial good. As the next chapter explains in detail, personal data is a concept granted human right protection.[91] Hence, its commodification and commercialisation can be seen as problematic *per se.*[92]

2.2.3. Infinite data storage

The 'datafication' of society and the advance of big data go hand in hand with the decreased costs of data management and data storage hardware and software. Exponential increases in computer capabilities, also referred to as Moore's law, created opportunities to build massive databases.[93] Today, data rendered in digital form can be stored for indefinitely long periods of time and can be readily retrieved.[94] To map and process large data sets, companies have developed powerful software such as Apache Hadoop integrating a MapReduce programming model.[95]

Databases are not necessarily stored locally but can be easily moved onto the Internet. Cloud computing describes a storing, processing, and use of data on remotely located computers accessed over the Internet.[96] Outsourcing data storage to an Internet service provider reduces cost and allows companies to collect data on a much larger scale.

The continuously changing Internet and rapid transformations of digital technologies create a perception of ephemerality of everything that happens online. The reality, however, is the opposite. Data shadows left behind individuals are increasingly difficult for those very individuals to shape, delete, or fully control. This is, in the first place, a consequence of the possibility of indefinite storage of personal information. However, the situation is escalated because data is typically moved and stored

---

writes about her doppelgänger – a digital representation of herself that commercial parties as well as the Government can reconstruct from the tracks she leaves on the internet: *'She is between the ages of 25–34. Or she's under 32. She is a millennial. She's inferred married. But she uses her phone like a single lady. She has eight lines of credit and is an upscale card holder. She's into coupons. She has recently purchased party goods, personal care products for men, and women's plus-sized apparel. She only walked 40,094 steps last week. She might qualify for a medical study on anorexia*.'

[89] Directorate General for Internal Policies, 'Big Data and Smart Devices and Their Impact on Privacy' (European Parliament 2015) 11.

[90] Zuboff, 77.

[91] See more in Chapter 3, section 3.2.

[92] European Data Protection Supervisor, 'Opinion 4/2017 on the Proposal for a Directive on Certain Aspects Concerning Contracts for the Supply of Digital Content' <https://edps.europa.eu/sites/edp/files/publication/17-03-14_opinion_digital_content_en.pdf> accessed 13 November 2017. In addition, problems may arise as a result of commodified data's secondary uses – see more in section 2.4.2. of this chapter that discusses some relevant risks.

[93] OECD, 'Exploring Data-Driven Innovation as a New Source of Growth' 10.

[94] Helen Nissenbaum, *Privacy in Context* (Stanford University Press 2010) 36.

[95] Kanala Urmila and Sandhya Rani, 'Hadoop Technology for BigData Analytics' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3168340> accessed 27 December 2018.

[96] European Commission, 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Unleashing the Potential of Cloud Computing in Europe' (2012).

in a cloud, where it is kept away from the physical reach of a user. Although cloud storage often proves convenient, e.g. when a user migrates her data from an old to a new device, it decreases the transparency of data processing and increases the safety risk as data is transferred over the Internet.[97]

2.2.4. Data analytics

As mentioned in the previous chapter, analytics is inherent to big data (data-driven) business models.[98] ENISA's definition of big data emphasises this, describing big data as *'the technologies, the set of tools, the data and the analytics used in processing large amount of data.'*[99] Data analytics is an umbrella term for various methods of information and knowledge extraction from large volumes of data to improve decision-making. Data mining is one of the analytical techniques used to analyse big data.[100] The goal of data mining is to discover previously unseen patterns and relationships from large datasets and to derive business value from these.[101] If these patterns or correlations are used to identify or represent people, they are referred to as *profiles.*[102] Analytical software offers several methods to perform data mining: statistical methods, e.g. regression and clustering, but also more sophisticated methods such as machine learning.[103] In machine learning, the machine automatically learns the parameters of models from the data using self-learning algorithms to improve its performance at a task with experience over time.[104] The discussion on data analytics is not complete without mentioning artificial intelligence (AI). AI incorporates machine learning and other disciplines such as robotics and natural language understanding; it is a broader concept of machines being able to carry out tasks in a way that would be considered 'smart'.[105] Today, AI is used in many contexts but most often to describe new, sophisticated technologies such as home assistants and autonomous weapons.

What makes all these big data techniques particularly valuable is the possibility of prediction.[106] Google's search engine, which uses anticipatory algorithms to predict what information users want based on a combination of data like website popularity, location, and prior search, is a prime example

---

[97] Christl writes about the use of Oracle's cloud, where data service providers upload their own data about customers, website visitors, or app users, combine it with data from many other companies and then transfer and utilize it on hundreds of other marketing and advertising technology platforms in real-time. This data can be used to, for example, find and target people across devices and platforms, personalize interactions, and eventually, to measure how consumers respond after having been addressed and affected on an individual level. Wolfie Christl, 'Corporate Surveillance in Everyday Life' *CrackedLabs* (June 2017) <http://crackedlabs.org/en/corporate-surveillance> accessed 27 May 2018.

[98] Chapter 1, section 1.1.

[99] ENISA, 'Big Data Security: Good Practices and Recommendations on the Security of Big Data Systems' (2015) 6.

[100] Herman T Tavani, 'KDD, Data Mining, and the Challenge for Normative Privacy' (1999) 1 Ethics and Information Technology 265. OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 144.

[101] Roger Brooks, 'Artificial Intelligence vs. Machine Learning vs. Data mining 101 - What's the Big Difference?' (*Guavas Blog*, 6 October 2017) <https://guavus.com/artificial-intelligence-vs-machine-learning-vs-data-mining-101-whats-big-difference/> accessed 14 June 2018.

[102] Bart HM Custers, *The Power of Knowledge: Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling in Epidemiology* (Wolf Legal Publishers 2004) 19.

[103] Galit Shmueli and others, *Data Mining for Business Analytics Concepts, Techniques, and Applications in R* (Wiley 2018) 20.

[104] Roger Brooks, 'Artificial Intelligence vs. Machine Learning vs. Data mining 101 - What's the Big Difference?' (Guavas Blog, 6 October 2017) <https://guavus.com/artificial-intelligence-vs-machine-learning-vs-data-mining-101-whats-big-difference/> accessed 14 June 2018. See more on machine learning in relation to profiling in Section 9.2.2.

[105] Bernard Marr, 'What Is The Difference Between Artificial Intelligence And Machine Learning?' *Forbes* (6 December 2016) <https://webcache.googleusercontent.com/search?q=cache:Qy4RIWTPArUJ:https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/+&cd=13&hl=en&ct=clnk&gl=us&client=safari> accessed 27 May 2018.

[106] Ian Kerr and Jessica Earle, 'Prediction, Preemption, Presumption: How Big Data Threatens Big Picture Privacy' (2013) 66 Stanford Law Review Online 65, 66.

of a 'big data prediction machine'.[107] Today, (predictive) data analytics occurs to a much greater extent and more easily compared to analytical endeavours in the past. This has been attributed to more extensive data gathering, the easier process of combining databases, and more powerful computer technologies to analyse the data.[108]

An example of sophisticated data analytics is ToyTalk, a US start-up company that operates the speech processing services for Hello Barbie and conducts analysis of the recordings of conversations between children and dolls.[109] ToyTalk's algorithm identifies sentences and phrases spoken aloud to the doll by converting them into text and analysing that text using the company's own application based on the knowledge gathered from Google Search, Wikipedia, and Weather Underground.[110] The technology enables the doll to respond to a child with lines of related, pre-recorded dialogue, adapted to every child's personal situation. With the help of analytics, Hello Barbie improves as a product, eventually increasing its value. However, abuses cannot be excluded. For instance, children's conversations could be mined to determine what products should be marketed to children and this information could be shared with advertisers.

2.3. How does the data-driven (big data) economy work?

The question critical to all commercial actors in the data economy is how to 'polish the data diamond' in the most profitable way. This leads to the topic of business models and strategies, which, in essence, describe ways in which companies use data to make money.[111]

This section aims to outline a high-level business model for data value creation to understand how the big data economy works and in what ways personal data is used. Of course, not all companies active on the data market follow the same strategy. Some of them are only involved in part of the process, and others choose their own strategy of value creation. However, for a typical data-driven business, the value creation model can be broken down into three key phases: 1) data collection, 2) data analytics and software, and 3) decision-making.

---

[107] Ibid., 67.
[108] Daniel J Solove, 'A Taxonomy of Privacy' (2006) 154 University of Pennsylvania Law Review 477, 506. Also see Helen Nissenbaum (2010) 43.
[109] <https://www.toytalk.com/about/> accessed 3 October 2016.
[110] Ibid.
[111] A business model is a term used to describe the strategy for data collection and reuse on such a large-scale. It can be explained through the value chain approach. In the internet economy, it typically consists of two parts of activities: first, activities associated with making something such as design, purchase of raw materials, manufacturing, and so on. Second part of the chain represent the activities associated with selling something: finding and reaching customers, transacting a sale, distributing the product or delivering the service. Joan Margaretta, 'Why Business Models Matter' *Harvard Business Review (May 2002).*

Figure 1: Data-driven value chain[112]

2.3.1. Data acquisition

The first issue that needs to be addressed is how data is generated and/or how it can be acquired. For obvious reasons, only examples of acquisition and generation of *personal* data are discussed here.[113] Although the list below is not exhaustive, it is broad enough to give an idea of the most common types of data collection and acquisition in the big data economy.

Data generating platforms such as Facebook, Google, and Strava are the most notorious collectors of personal data. These platforms generate data as a by-product of their actual business activity to support the sales of (digital) goods and services.[114] The main characteristic of service platforms is that they benefit from data enabling multi-sided markets.[115] On the one side of the market, platforms enter into relations with consumers by offering them free services; in exchange, they are able to capture a vast amount of these consumers' personal data. On the other side of the market, platforms contract with advertisers or other third parties that are willing to pay for the users' data captured by the platform.

Individuals' online activities are tracked beyond the activities of the data generating platforms. Features like IP addresses, authenticated logins, and cookies are exploited to monitor online behaviour.[116] Nowadays, almost every website is designed in a way that requires observing every visitor: time of visit, number of clicks, and moves across the screen. These activities represent only a part of all the personal data generated by the data economy but lie at the heart of many data-driven companies.[117]

Among indirect sources of personal data, *data brokers* are companies specialised exclusively in the provision of data. They compile personal data that comes from different suppliers and process it to enrich, clean, or analyse it.[118] The data is then provided to clients, such as social media and insurance companies. Typically, brokers' data is not sold but licensed.[119] Besides data brokers, data can also be

---

[112] Adapted from OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 132.
[113] Big amounts of non-personal data are generated as the by-product of an industrial activity. For instance, raw data collected by satellites or computer data generated in physics labs represents world's largest databases. Vivien Marx, 'The Big Challenges of Big Data' (2013).
[114] Cornelius Puschmann and Jean Burgess, 'The Politics of Twitter Data' in Katrin Weller and others (eds), *Twitter and Society* (Peter Lang Publishing, Inc 2014) 47.
[115] D. Daniel Sokol and Roisin Comerford, 'Antitrust and Regulating Big Data' [2016] Geo. Mason L. Rev. 1129, 1141.
[116] Nissenbaum (2010) 29.
[117] Ibid.
[118] Federal Trade Commission, 'Internet of Things: Privacy & Security in a Connected World' 3.
[119] <http://www.gartner.com/it-glossary/data-broker/> accessed on December 28, 2016.

acquired from commercial or non-commercial entities that generate data but are not willing or not able to reuse it in an innovative and profitable way. An example of data licensing (without any data broker being involved) is the agreement between the UK National Health Service (NHS) and Google DeepMind.[120] NHS owns a vast amount of medical data but has no capabilities to reuse it. Instead, it is interested in licensing or selling the data to someone with appropriate knowledge and technical capabilities. As a global leader in AI, Google is certainly a suitable partner.

Open data is gathered from publicly available records. Through open data initiatives, the public sector encourages access to and reuse of public data, including personal data. A recently founded public data distributor in the EU is the EU Open Data Portal, which provides access to large databases generated by EU institutions.[121] Open access to data is also offered by some private companies. Twitter, for instance, enables third parties to explore and reuse the data published on the platform.[122]

Machines and sensors that are part of the IoT generate a vast amount of data too.[123] Much of the data collected in the IoT environment relates, directly or indirectly, to individuals. As people are becoming more engaged with the technology, every aspect of their life is measured with sensors and analysed using big data analytics techniques. For example, radio-frequency identification (RFID) technology enables invisible monitoring of customers by tracking the tags of products that consumers put in their shopping carts.[124]

Finally, a *combination* of existing data sets is an additional way to acquire data. New data can be generated by analysing existing data, which in turn constitutes new personal data.[125] For example, after a user joins Facebook and consents to personal data processing, Facebook starts collecting vast amounts of her personal data. However, that is not all: the company combines these data sets with additional data purchased from data brokers to create a more precise picture of a user and to sell this enriched information to advertisers.[126]

From the perspective of an individual, data acquisition can be described as monitoring and tracking.[127] After all, every single visit to a website is registered by the website owner. Although visitors are typically asked to consent to the processing of personal data, tracking and monitoring can lead to subtle and disguised forms of data collection. These are rarely presented to the consenting individual in an informative and transparent way.

---

[120] Jane Wakefield, 'Google given access to London patient records for research' *BBC News* (3 May 2016) <http://bbc.com/news/technology-36191546> accessed 27 May 2018.
[121] <https://data.europa.eu/euodp/data/> accessed 27 May 2018.
[122] <https://developer.twitter.com/en/docs/tweets/search/overview> accessed 27 May 2018.
[123] See section 2.2.1.
[124] Article 29 Data Protection Working Party, 'Working Document on Data Protection Issues Related to RFID Technology' (2003).
[125] Manon Oostveen, 'Identifiability and the Applicability of Data Protection to Big Data' [2016] International Data Privacy Law 6.
[126] Evan Selinger and Brett Frischmann, 'Why it's dangerous to outsource our critical thinking to computers' *The Guardian* (10 November 2015) <https://www.theguardian.com/technology/2016/dec/10/google-facebook-critical-thinking-computers?CMP=Share_iOSApp_Other> accessed 28 December 2016. However, note that Facebook's CEO Mark Zuckerberg testified in 2018 that the company no longer uses this method of data enrichment. Transcript of the hearing of Mark Zuckerberg in the US Congress on April 10, 2018 <https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?utm_term=.013eea956ff1> accessed 28 May 2018.
[127] Also described as *dataveillance*. See for example David Lyon, 'Surveillance, Power and Everyday Life' in Robin Mansell and others (eds), *Oxford Handbook of Information and Communication Technologies* (Oxford Univeristy Press 2007).

When an individual is put at the centre, three categories of collected personal data can be distinguished:

(1) *Self-reported data,* or information that people volunteer about themselves, such as their email addresses, work and education history, and age and gender.[128] This often happens as part of a data generating platform activity. Examples include creating a social network profile and entering credit card information for online purchases.[129]

(2) *Digital exhaust,*[130] such as location data and browsing history, which is created when using mobile devices, web services, or other connected technologies.[131] In contrast to volunteered data, where the individual is actively and purposefully sharing his data, exhaust data can be generated even though a subject remains passive. This does not mean that digital exhaust is less useful or less revealing than self-reported data. In fact, the contrary can be true.[132]

(3) *Inferred data,*[133] or personal profiles used to make predictions about individual interests and behaviour, which are derived by combining self-reported data, digital exhaust, and other data.[134] An example is a consumer profile constructed by combining the RFID tags of the items purchased by a user and some other information about this specific consumer, e.g. her watch's RFID.[135] It is important to note that personal data can be also 'inferred' from pieces of 'anonymous' or 'non-personal' data.[136]

To determine how much consumers value their data, Harvard researchers examined the amount of money that survey participants would be willing to pay to protect different types of information.[137] The results showed that people valued self-reported data the least, digital exhaust more, and profiling data the most. Surprisingly, when it comes to legal protection of personal data, the order is not necessarily the same. Profiling data only recently received explicit and stronger protection under EU law.[138]

2.3.2. Data analytics and other software used to gain insights

The next step in the data value chain is the deployment of data analytics. From the business perspective, analytics is a segment of business intelligence that uses data tools to analyse and understand data. The task of analytics is to tailor data to draw useful actions and improve key performance indicators.[139]

---

[128] Timothy Morey, Theodore Forbath and Schoop Allison, 'Customer Data: Designing for Transparency and Trust' *Harvard Business Review* (May 2015).
[129] OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being'.
[130] In the OECD terminology, this is *observed* data. I use the expression *digital exhaust* as it appears more illustrative.
[131] Morey, Forbath and Allison (2015).
[132] For example as metadata - data that provides information about other data. Merriam-Webster online dictionary <https://www.merriam-webster.com/dictionary/metadata> accessed 28 December 2016.
[133] Used by the OECD and Mireille Hildebrandt. See OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being'; Hildebrandt (2008).
[134] Morey, Forbath and Allison (2015).
[135] Article 29 Data Protection Working Party, 'Working Document on Data Protection Issues Related to RFID Technology' 6.
[136] OECD 'Data-Driven Innovation: Big Data for Growth and Well-Being' 152.
[137] Morey, Forbath and Allison (2015).
[138] In the upcoming GDPR, profiling data is explicitly protected under Article 20. See Chapter 9 for more details.
[139] Mayer-Schönberger and Cukier (2014).

The OECD distinguishes three main functions through which companies use data analytics to gain insights:[140]

(1) *Extracting information from unstructured data*. Unstructured data is by far the most frequent type of data but is hardly useful. To structure it in a predefined data model, different analytical techniques can be applied.[141] Modern analytics enables insights into databases that were not possible in the past. For example, a large collection of photos can be interpreted by using analytical software. One such well-known photo recognition algorithm has been developed by Facebook. This technology has given Facebook *'the ability, in a semantically appropriate way, to describe what's happening in a photo, that's very advanced and starting to approach the holy grail of image recognition.'*[142]

(2) *Digital real-time monitoring*. The fact that data is collected at a high speed and can be processed and analysed instantly represents a large benefit for the economy. By gaining real-time insights, companies are able to base decisions on evidence that is very close to the actual market situation (e.g., customers' current preferences, trends). Such monitoring may also benefit consumers. An illustrative example comes from the energy sector. BC Hydro is an electric utility providing power to nearly 2 million Canadian residents.[143] In 2011, the company began upgrading its electricity meters to smart meters. Users can now track their energy use per hour and see trends in their own usage data. As a result, it is easier for them to keep their use of energy and spending under control.

(3) Inference and prediction. Sophisticated data analytics supports data-driven inference and prediction. The discovery of knowledge is now possible even if there was no prior record of such information.[144] To use the example above, Facebook's face recognition algorithm is not only able to organise users' photos but also to automatically identify and tag users according to user-provided photos. Netflix, a data-driven company that collects information from its 50 million plus subscribers at an extraordinary speed, is known for its sophisticated predictive algorithm which is able to provide personalised movie suggestions to individual users.[145]

Appreciating the power of information to analyse people and to predict and even control their actions is nothing new. In fact, understanding and foreseeing human behaviour has always been a part of human social relations and interaction.[146] However, for the reasons explained above, it now occurs to a much greater extent. In fact, analytics represents an increasingly important aspect of the modern data economy. Some of the world's most successful and innovative companies, such as Google,

---

140 OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 150-153.
141 See section 2.2.4. for some examples of analytical techniques.
142 <http://digiday.com/platforms/facebooks-new-image-recognition-technology-data-windfall-advertisers/> accessed on December 28, 2016.
143 Conner Forrest, 'Ten Examples of Iot and Big Data Working Well Together' *ZDNet* (2 March 2015) <http://www.zdnet.com/article/ten-examples-of-iot-and-big-data-working-well-together/> accessed 27 May 2018.
144 OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 36.
145 A Porat and LJ Strahilevitz, 'Personalizing Default Rules and Disclosure With Big Data' (2014) 112 Michigan Law Review 1417, 1451-1452.
146 Nissenbaum (2010) 43.

Facebook, Amazon, and eBay, have built their business model on the analytical exploitation of big data.[147]

In general, analysts are interested in trends, models, and correlations, and not in a specific individual. However, individuals can be greatly affected by the failures of the analytical processes when applied to them, e.g. the use of biased data.[148] Consider the following example. A company wants to create the ideal profile for its next top manager. On the basis of the available data, the algorithms discover that the ideal top manager is a middle-age white male. As the database probably contained many top managers with this profile, the resulting pattern only confirms the discriminatory tendency in the hiring processes.[149] These negative consequences escalate in the final stage of the data value chain, when data-driven decisions are made. Section 2.4.2. explores those risks in more detail.

2.3.3. Generating value through decision-making

The final step in the data value chain is acting upon discovered knowledge, i.e. using insights in data to draw useful decisions that generate value. This knowledge can either be a result of the analysis of a company's own data or it can be derived from a third party's data. The process of decision-making can be supervised by a human, but it can also run autonomously without any human interference.

Netflix, an American streaming media provider, serves as a model for decision-making based on internal data analysis. As mentioned above, Netflix uses sophisticated predictive algorithms that are able to provide personalised movie suggestions to individual users. This specific knowledge of users' movie preferences also drives the company's business decisions. By observing that subscribers who watched the original British version of House of Cards were highly likely to watch movies starring Kevin Spacey or directed by David Fincher, the company predicted the success of the new House of Cards series. Eventually, the company started licensing the series, which was a great success and brought it a large profit.[150]

Decision-making that is based on multiple input is typical to the field of behavioural targeting. An advertising network follows an Internet user's behaviour while he surfs the Web. Different technologies enable the tracking of consumers, but typically *'cookies are used […] to identify users who share a particular interest …'*.[151] By knowing users' shopping preferences, the network is able to create customer profiles and provide each user with an individually targeted advertisement. Ad networks' insights are utilised by advertisers to target consumers with more relevant ads.

An example of a more complex yet promising strategy of profit generation is the agreement between Google and the NHS. The agreement has allowed Google to access about 1.6 million patient records.[152] Google's AI division DeepMind will use the data to develop an early warning system for patients at risk

---

[147] Fortune 500 list for 2016 shows the world's most successful companies. Data-driven companies top the list – Walmart is on the first and Apple is on the third place <http://fortune.com/fortune500/> accessed 24 January 2016.
[148] Oostveen (2016) 7.
[149] Bart Custers and Helena Ursic, 'Worker Privacy in a Digitalized World under European Law' 39 Comparative Labor Law & Policy Journal 323.
[150] OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 152-153.
[151] Frederik Zuiderveen Borgesius, 'Singling out People without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation' (2016) 32 Computer Law & Security Review 5.
[152] Jane Wakefield, 'Google given access to London patient records for research' *BBC News* (3 May 2016) <http://www.bbc.com/news/technology-36191546> accessed 27 May 2018.

of developing acute kidney injuries.[153] Insights will be also used by hospitals to streamline and improve health treatments.[154]

The idea of Google as the NHS health data user has triggered much public disapproval for reasons concerning possible interference with collective and individual rights. Some have criticised the fact that a vast amount of data was transferred to Google with such ease, and pointed at the long-lasting relevance of knowledge that might be hidden within the set.[155] That discussion partly explains why the third step in the data value creation model matters from the individual perspective. However, data-driven decisions also have some short-term consequences. For example, Netflix has mastered the ability to identify and recommend movies that keep us attached to its service.

What is also relevant from an individual point of view is the automatisation of decision-making. More and more decisions are made without any human involvement.[156] While automatisation has a great potential for the economy, it also opens some difficult issues related to the protection of individuals. Some concerns regarding non-transparency and power asymmetries are explored in sections 2.4.2.2. and 2.4.2.4.

2.4. The individual in the data-driven economy

A vast amount of information in the digital universe is created by *individuals,* including phone calls, emails, photos, online banking transactions, and postings on social networking sites such as Twitter.[157] Moreover, data generated by machines or combined from several sources can, directly or indirectly, be related to an individual person.[158]

Every online appearance leaves a digital trace, which gradually grows into a vast registry of the actions constituting 'data doubles' or 'quantified selves'.[159] By participating in these activities, individuals actively contribute to and co-create the data economy. As shown above, this data represents the raw material for the entire online service industry.

The fact that personal data sources are highly interesting and useful in the data economy has led to their commodification and commercialisation. Commodified personal data is a discrete package of personal information that can be exchanged for something else.[160] Using data as an object in a business exchange creates (monetary) value. However, by taking the form of a commodity, personal data is devaluated to the level of a commercial good. For a concept that has been granted human rights protection, this is a delicate transformation.[161]

---

153 Ibid.
154 Ibid.
155 Julia Powles and Hal Hodson, 'Google DeepMind and Healthcare in an Age of Algorithms' (2017) 7 Health and Technology 351.
156 OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being' 145-155.
157 Conner Forrest, 'Ten Examples of Iot and Big Data Working Well Together' *ZDNet* (2 March 2015) <http://www.zdnet.com/article/ten-examples-of-iot-and-big-data-working-well-together/> accessed 27 May 2018.
158 Sandra Wachter, 'Normative Challenges of Identification in the Internet of Things: Privacy, Profiling, Discrimination, and the GDPR' (2017) 14 <https://ssrn.com/abstract=3083554> accessed 28 May 2018.
159 Gemma Galdon Clavell, 'Policing, Big Data and the Commodification of Security' in Bart Van Der Sloot, Dennis Broeders and Erik Schrijvers (eds), *Exploring the Boundaries of Big Data* (Amsterdam University Press 2016) 106.
160 Paul Schwartz, 'Property, Privacy and Personal Data' (2003) 117 Harvard Law Review 2056, 2069.
161 See Chapter 3 on data protection law and its constitutional roots.

In estimating the impacts of the data economy on individuals, authors in the economic field seem to contradict each other. Some believe that data can play a significant economic role to the benefit of both private commerce and national economies and see a great benefit of the data-driven economy for the well-being of citizens.[162]

In contrast, others warn that an expanding consumer surplus means that many producer surpluses have been competed away.[163] In other words, the data-driven economy largely benefits individuals and leaves commercial actors worse off. However, a third stream of economists believe that the big data economy in fact decreases consumer surplus.[164] Their argument is in line with those who question data economy benefits due to decreasing privacy protection, discrimination, and other negative implications.[165]

Preceding sections have already indicated that data exchange carried out on the Internet is not only innocent and advantageous. Several risks of data-driven business processes have already been briefly mentioned: non-transparent data tracking and monitoring, decreased control over data stored in the cloud, and inexplicability of algorithms.

As the data-driven economy is becoming more influential, gains and opportunities need to be carefully balanced. This requires a trade-off between big data risks and rewards. In fact, striking the right balance could be one of the greatest public policy challenges of our time.[166]

2.4.1. Benefits

Taking part in the data economy has the potential to enhance an individual's overall well-being. The term well-being is understood as a combination of social, economic, and psychological states associated with positive feelings.[167] Four groups of benefits that enhance well-being are described shortly below. The list is not exhaustive because it only clusters those that are most commonly observed. Not only individual benefits but also positive outcomes of data processing for society as a whole are taken into account.

2.4.1.1. Convenience

Sharing personal data can be convenient. For example, most people prefer to use a credit card rather than a debit card for the convenience of a deferred payment, although this eventually decreases the confidentiality of their purchases.[168] In a similar vein, when someone agrees to share data with advertisers, she is alerted about relevant offers and redirected to the most relevant product.

---

[162] McKinsey, 'Big Data: The next Frontier for Innovation, Competition, and Productivity' 1-2. Also see OECD, 'Data-Driven Innovation: Big Data for Growth and Well-Being'.
[163] Brian Kahin, 'Digitization and the Digital Economy' (2013) <http://ssrn.com/abstract=2782906> accessed 27 May 2018.
[164] Anna Bernasek and DT Mongan, *All You Can Pay* (Nation Books 2015).
[165] See for example Tal Z Zarsky, '"Mine Your Own Business!": Making The Case For The Implications Of The Data Mining Of Personal Information In The Forum Of Public Opinion' (2002) 5 Yale Journal of Law and Technology 1306.
[166] Lynskey (2015) 202.
[167] For the definition of well-being see Carol D Ryff 'Happiness is everything, or is it? Explorations on the meaning of psychological well-being' [1989] 57 Journal of personality and social psychology 1069.
[168] Kent Walker, 'Where Everybody Knows Your Name: A Pragmatic Look at the Costs of Privacy and the Benefits of Information Exchange' (2000) 190 Stanford Technology Law Review 1.

Withholding contact information typically results in limited availability (or unavailability) of the discounts and offers such as free videos, discounts on children's toys, or cut-rate airfares.[169]

2.4.1.2. Self-expression and self-control

With recent technological advances, personal data may also become a way of self-expression leading to one's own (representation of) identity.[170] For instance, the Strava app has created a large 'quantified self-movement' community.[171] Users who track their sport activities are encouraged to share and analyse results on the platform. Large-scale data collection and real-time aggregation of data give them an opportunity to monitor their training, compare it against that of other peers, and challenge themselves by setting higher goals.

The quantified-self apps often offer functions beyond fitness tracking. For instance, women trying to conceive use apps to track their periods, basal temperature, weight, mood, and sex life.[172] If they feed the app with enough data, the algorithm is able to calculate their ovulation date and assess their chance of pregnancy.

2.4.1.3. Reduced cost and/or (in)direct monetary benefits

The use of personal data reduces marketing and distribution costs for both businesses and consumers, and thus ultimately decreases the prices of all goods and services.[173] Targeted offers help sellers avoid investing time and resources in targeting uninterested buyers, which translates to a better price for those who do actually show interest in the products.

SkyScanner's search algorithm proves that a company's successful data-driven service can directly benefit consumers.[174] SkyScanner is a global metasearch engine for information on the World Wide Web that enables people to find comparisons for flights, hotels, and car hire services. It gathers data from numerous airliners and comes up with a selection of the most affordable flights based on the user's preferences. Those who use the service agree that it makes life much easier and are reluctant to abandon it.[175]

The technological revolution driven by big data has the potential to empower individuals even more profoundly. Since Web 2.0 emerged, users have been given exciting opportunities to take a more active role on the Internet.[176] Today, consumers no longer passively observe the online market, but actively engage in the economic exchange. An empowered and more independent consumer can instantly move between the two poles of consumption and production. Airbnb is a platform which, with the help of big data, connects property owners with those searching for short-term accommodation. By using Airbnb's big-data-driven service, an owner can effortlessly transform into a quasi-commercial

---

169 Ibid.
170 Michiel Rhoen, 'Beyond Consent: Improving Data Protection through Consumer Protection Law' (2016) 5 Internet Policy Review.
171 <https://www.strava.com> accessed 27 May 2018.
172 <https://itunes.apple.com/us/app/glow-ovulation-period-tracker/id638021335?mt=8> accessed 27 May 2018.
173 Walker (2000) 8.
174 Rubinstein (2013).
175 Skyscanner was one of Scotland's most highly valued start-ups. The company has grown exponentially since it was set up in 2003. Steve Vance, 'Skyscanner acquired by Chinese travel giant Ctrip in a £1.4 billion deal' *Citya.m.* (19 September 2016) <http://www.cityam.com/248684/significant-investment-puts-skyscanner-firmly-position> accessed 27 May 2018.
176 Daly (2016).

party with access to the global market. Not only does this reduce the owners' cost, it also opens up possibilities for direct monetary benefits.

Furthermore, some assert that individuals should be free to derive some direct benefit, including monetary, from the use of their personal data.[177] Personal datasets could be licensed to third parties in exchange for additional services, e.g. free social networking, or for cash value.[178] In some sense, this is already happening. The notions of economic value and ownership of personal data are a reality of modern data processing practices.[179] However, full application of the property law regime to personal data[180] is probably not feasible under the current legislation,[181] nor is it in line with the European fundamental rights doctrine, which perceives protection of data as an unalienable right.[182]

2.4.1.4. New knowledge and innovations

Data collected via various media – Internet, communication, cameras – works as an asset and raw material for commercial actors in the data economy and for science and society at large. Patterns of behaviour identified on the group level can have long-reaching consequences. In the healthcare sector, platforms constructed based on data from millions of patients and their health records have the potential to revolutionise clinical research and to bring significant benefits to many stakeholders, including patients, health systems researchers, industry, and society.[183] An example is the Dutch start-up Filterless, which has developed software to combine millions of health data (mostly genome) collected by hospitals, insurance companies, and pharmaceutical companies, and mine it with the goal of finding new insights that could help healthcare providers improve treatments for all patients.[184]

Two conditions must be met to generate data-driven insights. First, individuals have to share their data. The technology's value increases with the number of people who use it or permit their information to be shared. Sharing customer data on a large scale creates a consumer community where each participant can benefit from the experience and information of a fellow customer. For instance, Amazon recommends additional books that users might like based on the purchasing patterns of others who have bought the same books in the past.[185] Had all fellow users refused to provide a recommendation, assessing a book's suitability would become much more difficult and costly. Walker contends that certain socially beneficial products and services can only exist if everyone agrees to participate, and calls the opposite situation a 'tragedy of commons' that can best be illustrated with a phone directory: unless the majority of people agree to share information, the directory is useless.[186]

---

[177] See for example Lessig (2006) 228.

[178] European Data Protection Supervisor, 'Privacy and Competitiveness in the Age of Big Data: The Interplay between Data Protection, Competition Law and Consumer Protection in the Digital Economy' (2014).

[179] Nadezhda Purtova, 'The Illusion of Personal Data as No One 's Property' (2015) 7 Law, Innovation and Technology 83, 4.

[180] E.g. using a license to perform an infinite transfer of personal data from an individual to a commercial entity.

[181] See for example EU Charter, Article 8.

[182] See for example: European Court of Human Rights, Sanles Sanles v. Spain, app.no. 48335/99; European Court of Human Rights, Thévenon v. France, app.no. 2476/02; European Court of Human Rights, Mitev v. Bulgaria, app.no. 42758/07; European Court of Human Rights, M.P. and Others v. Bulgaria, app.no. 22457/08; European Court of Human Rights, Koch v. Germany, app.no. 497/09.

[183] P Coorevits and others, 'Electronic Health Records: New Opportunities for Clinical Research' (2013) 274 Journal of Internal Medicine 547.

[184] <http://filterless.nl/index.php/drug-discovery/> accessed 27 May 2018.

[185] Walker (2000) 14.

[186] Ibid., 12.

The second condition for knowledge generation and innovation is openness of data. This is why policy-makers strongly encourage governmental data sharing. The objective is to reuse data and accrue value for businesses and for citizens. One good practice comes from Chicago, where the municipal open data platform supported app developers to build innovative solutions based on public data. For example, developers created an interactive map that lets citizens find out how a building is zoned, learn where to locate a business, or explore zoning patterns throughout the city.[187] Due to many positive side effects such as transparency, increased trust, and added value, EU member states are required to make as much public information available for reuse as possible.[188]

2.4.1.5. Security of data and citizens

With more data available and processed, it is growing increasingly difficult to disguise someone's identity. This in turn means better security, as fraudulent individuals and entities can no longer take part in the digital market.[189] Furthermore, being able to assess vast amounts of data to identify suspicious incidents can help identify criminal behaviour and prevent costs. Palantir, a US software company, has been selling its big data tools to the police and government departments to flag traffic offenses, parole violations, and other everyday infractions.[190] Palantir's software can ingest and sift through millions of digital records across multiple jurisdictions, spotting links and sharing data to make or break cases. The police have neither the technical resources nor sufficient data to carry on such an analysis themselves.[191] For better or worse, the police departments that deploy Palantir have become dependent upon it for some of their most sensitive work.

2.4.2. Risks

Big data is a recent, evolving phenomenon. As a consequence, many of its risks are yet to be explored. In addition, because big data activities are often carried out in disguise and are highly complex, the risks are difficult to notice.

Therefore, it is impossible to provide an exhaustive list of big data risks. Instead, the analysis should focus on a limited number of core values that can be compromised as a result of big data business practices. To tackle this task, the following sections build on Richards and King's framework of three paradoxes of big data: the transparency, the identity and the power paradox.[192] These paradoxes elucidate the values that can be undermined by the growing big data economy: privacy, transparency, autonomy, and power symmetry.

---

187 <https://secondcityzoning.org> accessed 27 May 2018.

188 Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information [2013] OJ L175/1.

189 The danger is, however, that the state's concerns for security can be turned into undesirable surveillance. For example, China recently introduced a social credit scoring system based on citizens' behavior on social networks. See for example Rachel Botsman, 'Big data meets Big Brother as China moves to rate its citizens' (*21 October 2017*) <http://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion> accessed 27 May 2018.

190 Mark Harris, 'How Peter Thiel's Secretive Data Company Pushed Into Policing' *Wired* (8 September 2017) <https://www.wired.com/story/how-peter-thiels-secretive-data-company-pushed-into-policing/> accessed 27 May 2018.

191 Ibid.

192 Neil M Richards and Jonathan J King, 'Three Paradoxes of Big Data' (2013) 66 Stan. L. Rev. Online.

2.4.2.1. Compromised privacy

Privacy is a concept that allows for multiple definitions.[193] Chapter 3 of this thesis provides a detailed analysis of the term and traces back attempts to capture its meaning. For now, it suffices to understand privacy in its ordinary sense: as an attribute of things that affect or belong to private individuals, that are generally distinct from the public, and that are kept confidential and secret (e.g. not disclosed to others and kept from public knowledge or observation).[194]

The data-driven economy often gives the impression that privacy has been eliminated or is even dead.[195] Mark Zuckerberg, Facebook's CEO, argued that privacy has fundamentally evolved in recent years and can no longer be seen as a social norm.[196] While it is true that privacy as a social norm has been transformed, it has not lost any of its strength. On the contrary, considering the many new types of privacy violations, some of which are mentioned below, privacy has never been more relevant. Zuckerberg himself is proof. In a photo shared via Twitter in the summer of 2016, his computer can be seen, on which the camera and headphone jack are covered with tape, and the email client he uses is Thunderbird (a popular email client among the tech-savvy, and particularly those who want to use PGP-encrypted emails).[197] Zuckerberg's example may sound anecdotal but it is an indicator of a wider trend, suggesting that people increasingly care about keeping his work and conversations private.

In the data-driven economy, dataveillance is what most apparently puts privacy at risk. Dataveillance is the systematic use of personal data systems in the investigation or monitoring of the actions or communications of one or more persons.[198] In the data economy, in which individuals' behaviour and all their actions are increasingly datafied, dataveillance is easy to conduct. Clarke observes that it is significantly less expensive than physical and electronic surveillance, because it can be automated. As a result, the economic constraints on dataveillance are diminished, and more individuals, and larger populations, can be monitored.[199] Dataveillance can be particularly dangerous because it enables the inference of facts that someone would rather keep secret. For example, a person shares information about her hobbies or favourite books but not information about her sexual orientation. However, by using big data techniques, this information can be predicted anyway. Kosinski, Stillwell, and Graepel have shown how a range of highly sensitive personal characteristics, including sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, and parental separation, can be predicted highly accurately on the basis of Facebook likes.[200] Connected devices are another example of how big data quickly escalates into riskier conduct. IoT enables easy integration, aggregation, or correlation of various aspects of users' identity.[201] At first glance, this gathering of data is just a step in the data-driven value chain (data

---

[193] Daniel J Solove, 'Conceptualizing Privacy' (2002) 90 California Law Review 1087, 1088.
[194] The Concise Oxford dictionary (1990), Black's Law Dictionary (1910).
[195] Neil M Richards and Jonathan J King, 'Big Data Ethics' (2014) 49 Wake Forest Law Review 393, 409.
[196] Bobbie Johnson, 'Privacy no longer a social norm, says Facebook founder' (10 January 2010) <https://www.theguardian.com/technology/2010/jan/11/facebook-privacy> accessed 27 May 2018.
[197] Katie Rogers, 'Mark Zuckerberg Cover His Laptop Camera. You Should Consider It, Too.' *The New York Times* (22 June 2016) <https://www.nytimes.com/2016/06/23/technology/personaltech/mark-zuckerberg-covers-his-laptop-camera-you-should-consider-it-too.html> accessed 14 June 2018.
[198] <http://www.rogerclarke.com/DV/Intro.html#Priv> accessed 3 October 2016.
[199] <http://www.rogerclarke.com/DV/Intro.html#Priv> accessed 3 October 2016.
[200] M Kosinski, D Stillwell and T Graepel, 'Private Traits and Attributes Are Predictable from Digital Records of Human Behavior' (2013) 110 Proc Natl Acad Sci USA 5802.
[201] Kord Davis, *Ethics of Big Data* (O'Reilly Media, Inc 2012) 33.

acquisition). However, it may lead to disclosure of some highly sensitive details.[202] For instance, electronic toothbrushes may reveal how often and how long people brush their teeth, indicating potential health issues.[203]

In the big data economy, even anonymised data cannot guarantee privacy. In fact, anonymised data can be as useful as personal data in many cases.[204] A typical example is a company that wants to personalise its marketing campaigns with the help of profiling. The use of personal data may be helpful to assess which people are potentially interested in particular products or services, but aggregated data on the street or neighbourhood level may be similarly useful and cheaper to process. Inferring information from group profiles supports predictions about someone's personal circumstances. As soon as '*[t]hree or four data points of a specific person match inferred data (a profile), which need not be personal data […]',*[205] a company is able to highly accurately predict characteristics of individual users.[206]

The flow of data among the actors in the data-driven economy escalates the risk of privacy intrusions. This is why Nissenbaum believes that meeting individual expectations about the flow of personal information sits at the core of privacy.[207] The section on data acquisition mentioned a number of data sources, including data brokers. Specifically, it pointed to the fact that data is often acquired by means of data combination. For example, Facebook's own databases are merged with detailed dossiers obtained from commercial data brokers about users' offline life.[208] In this way, Facebook improves its own data with categories that users did not share or want to reveal on Facebook. If information is used in contexts that are at odds with individuals' expectations, this can lead to feelings of awkwardness and discomfort.[209]

2.4.2.2. Lack of transparency

Transparency describes something that is easy to perceive or detect, and is open to scrutiny. In contrast, non-transparency can be illustrated with the metaphor of a black box: a complex system or device whose internal workings are hidden or not readily understood.[210] In the context of data-driven decision-making, the black box metaphor stands for outcomes that emerge without satisfactory explanation.

Transparency is the second value at risk in the era of the data-driven economy. Although big data promises to make the world more transparent, its collection is invisible and its tools and techniques

---

202 Custers and Ursic, 'Worker Privacy in a Digitalized World under European Law' 330.
203 Ibid.
204 Daniel Bachlechner and others, 'WP1 Mapping the Scene: D1.2 Report on the Analysis of Framework Conditions (Deliverable for the EuDEco H2020 Project)' (2015) 30 <https://www.universiteitleiden.nl/binaries/content/assets/rechtsgeleerdheid/instituut-voor-metajuridica/d1.2_analysisofframeworkconditions-v1_2015-08-31-1.pdf> accessed 14 June 2018.
205 Hildebrandt (2013) 33.
206 Porat and Strahilevitz (2014) 1440.
207 Nissenbaum (2010).
208 Julia Angwin, Terry Parris Jr. and Surya Mattu, 'Facebook is quietly buying information from data brokers about its users' offline lives' *Business Insider* (30 December 2016) <http://www.businessinsider.com/facebook-data-brokers-2016-12?r=UK&IR=T> accessed 14 June 2018.
209 Nissenbaum (2010) 21.
210 <https://en.oxforddictionaries.com/definition/black_box> accessed on 9 January 2017.

opaque, curtained off by layers of physical, legal, and technical protection.[211] Non-transparent processing of data occurs in all three stages of the data-driven value chain: when data is acquired, when it is analysed, and when it is used. To illustrate the problem, three examples are given below: privacy policies, algorithmic black box, and the cloud computing black box.

Privacy policies (notices). The ubiquitous and automated collection of data in the data-driven economy is by definition opaque. Law requires data collectors to draft privacy policies to explain in what ways and under what circumstances personal data is collected, used, and shared. However, it has been shown that the objectives of privacy policies are flawed as people cannot understand the complex legalistic language, and policies are not specific enough to plausibly present what an individual actually consents to.[212]

Algorithmic black box. To extract useful patterns and create profiles, enormous amounts of consumer data are mined using complex algorithms.[213] As Pasquale points out, the key problem is that little is known about data mining and subsequent choice architecture processes.[214] In spite of being increasingly used to derive all sorts of findings, hidden algorithms are shrouded in secrecy and complexity. Barely anyone is able to fully capture how algorithms work and to monitor their actions. For example, Acxiom, the online data marketplace, is used as a source of numerous data points for an algorithm to determine a customer's creditworthiness.[215] Because of a bad credit score calculated on the basis of aggregated information, a consumer will be charged more, but she will never understand how exactly this amount was calculated or know what information Acxiom provided.[216] In addition, not even engineers working with the algorithms are fully able to capture their nature and monitor their actions.[217]

Cloud computing black box. The black box problem is duplicated in the cloud computing environment, mainly due to indefinite and non-transparent storage. In most cases individuals are unaware of what actually occurs in a cloud. Data can be shared with third parties, sold to advertisers, or handed over to the government. The loss of transparency on the Internet results in the feeling of powerlessness. As Schneier puts it, *'trust is our only option. There are no consistent or predictable rules. We have no control over the actions of these companies. I can't negotiate the rules regarding when yahoo will access my photos on Flicker. I can't demand greater security for my presentations on Prezi or my task list on Trello. I don't even know the cloud providers to whom those companies have outsourced their infrastructures […]. And if I decide to abandon those services, chances are I can't easily take my data with me.'*[218]

[211] Richards and King (2013) 42.
[212] Simone van der Hof, Bart W Schermer and Bart HM Custers, 'Privacy Expectations of Social Media Users: The Role of Informed Consent in Privacy Policies' (2014) 6 Policy and Internet 11.
[213] On the definition of profiling see Custers (2014) 156.
[214] Frank Pasquale, *The Black Box Society* (Harvard University Press 2015).
[215] See for example Mikella Hurley and Julius Adebayo, 'Credit Scoring in the Era of Big Data' (2016) 18 Yale Journal of Law and Technology 148, 175.
[216] Ibid.
[217] '*... even those on the inside can't control the effects of their algorithms. As a software engineer at Google, I spent years looking at the problem from within …*' David Auerbach, 'The Code We Can't Control' *Slate* (14 January 2015) <http://www.slate.com/articles/technology/bitwise/2015/01/black_box_society_by_frank_pasquale_a_chilling_vision_of_how_big_data_has.html> accessed 27 May 2018.
[218] Bruce Schneier, *Data and Goliath* (WWNorton & Company 2015) 115.

2.4.2.3. Undermined autonomy

Faden and Beauchamp define autonomy in practical terms as *'the personal rule of the self by adequate understanding, while remaining free from controlling interferences by others and from personal limitations that prevent choice.'*[219] Three dimensions of autonomy stem from this definition: self-governance (control), freedom from interference of others, and free choice. The examples below show how big data undermines each of them.

Free choice can be restricted as a result of limited confidentiality and privacy of personal data traces on the Internet. Knowing that the US National Security Agency (NSA) can follow every move we make might deter us from using a US online service.[220] The abstention from an action or behaviour due to the feeling of being observed is described as a *chilling effect*.[221] However, in some cases, the feeling of being watched creates a *nudge* for individuals to act. For example, research has shown that people pay more for coffee on the honour system[222] if eyes are depicted over the collection box.[223] Individuals' attitudes and behaviours change in such circumstances even though no real person is there. In 2009, German politician Malte Spitz sued a mobile network operator to obtain access to his mobile phone data. He then passed the information to Zeit Online, a news website, which created a visualisation showing where he had been and what he had been doing.[224] His profile was strikingly detailed, showing when Spitz had walked down the street, when he had taken a train, when he had been in an airplane, and where he had been in the cities he had visited. It also showed when he had worked, when he had slept, when he could have been reached by phone and when he had been unavailable, when he had preferred to talk on his phone, and when he had preferred to send a text message. It even showed which beer gardens he had visited in his free time.[225] Let us assume for the moment that these visits were not just frequent but rather excessive. If politician Spitz had known that someone could track his weekend visits to the beer gardens, would he still have been such a regular visitor, or would he have chosen a more neutral spot to spend his free time?

Another example of compromised autonomy is linked to non-transparent data processing and decision-making. In 2009, Eli Pariser noted that the news he received and search results that appeared on Google differed substantially from those viewed by his colleagues.[226] He soon realised that the reason was his personalised news website. Namely, based on his user profile and corresponding group profiles, the website was able to learn about his inferred political interests, which in turn meant that it could give more prominence to his favourite political group's media items. He described the situation

---

[219] Quoted in: Bart W Schermer, Bart Custers and Simone van der Hof, 'The Crisis of Consent' [2013] Ethics & Information Technology 6.
[220] Glenn Greenwald, *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State* (Metropolitan Books, 2014).
[221] Jonathon W Penney, 'Chilling Effects : Online Surveillance and Wikipedia Use' (2016) 31 Berkeley Technology Law Journal.
[222] A system of payment or examinations which relies solely on the honesty of those concerned.
[223] Ryan M Calo, 'The Boundaries of Privacy Harm' (2011) 86 Indiana Law Journal 1131, 1147.
[224] Dan Smith, 'Why can't we see the personal data we produce?' *The Telegraph* (5 July 2013) <http://www.telegraph.co.uk/sponsored/technology/technology-trends/10161697/personal-data.html> accessed 28 May 2018.
[225] Kai Biermann, 'Betrayed by our own data', *Zeit Online* (10 March 2011) <http://www.zeit.de/digital/datenschutz/2011-03/data-protection-malte-spitz> accessed May 28 2018.
[226] Eli Pariser, 'Beware Online Filter Bubbles' *TedX Talk* (March 2011) <https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles?language=en> accessed 3 October 2016.

as a *filter bubble*: *'a synonym for a unique universe of information for each of us.'*[227] The filter bubble represents the risk of seriously limiting someone's free choice. For example, when users of such personalised services form their political ideas, they may encounter fewer opinions or political arguments.[228]

2.4.2.4. Power asymmetries

In the data-driven economy, power is linked to two dimensions: 1) the access to data and control over it, and 2) the ability of sophisticated data processing.[229] The power asymmetry is most apparent in the relationship between data-driven businesses and individuals. However, it can also be observed in relationships between other actors in the economy. Small businesses often become dependent on and powerless in relation to big data holders.[230] Finally, power asymmetry affects authorities too, as they struggle to understand the data-driven economy and its consequences. '*[T]o understand what is going on we have to go for geeks,'* stated the director of the European Consumer Organisation to express her frustration with the data economy black box.[231]

To a large extent, the asymmetry between data controllers and individuals stems from the architecture of the data-collecting platforms. Because of these platforms' design, it is easy for them to take full ownership of users' input, e.g. photos, comments, and texts. In such circumstances, users' control over data fades away. Until recently, users of the dating app Tinder were asked to give away control of their pictures, videos, and chat logs forever.[232] Although individuals certainly benefit from the digital economy, e.g. by being able to use the Amazon online shopping tool, they pay (often unknowingly) for these services with their non-monetary assets, and their input is not always fairly evaluated.[233] In addition, the architecture of the platforms disables transparency. As explained above, algorithms that drive the functioning of the platforms are shrouded in secrecy and complexity, and barely anyone is able to fully capture how they work.

---

[227] Ibid.

[228] Filter bubble could even interfere with collective goods such as democracy. Harvard Law professor Jonathan Zittrain explained in 2010 how *'Facebook could decide an election without anyone ever finding out',* after the tech giant secretly conducted a test in which they were able to allegedly increase voter turnout by 340,000 votes around the country on election day simply by showing users a photo of someone they knew saying "I voted". Trevor Timm, 'You may hate Donald Trump. But do you want Facebook to rig the election against him?' *The Guardian* (19 April 2016) <https://www.theguardian.com/commentisfree/2016/apr/19/donald-trump-facebook-election-manipulate-behavior> accessed 28 May 2018. See also Robert M Bond and others, 'A 61-Million-Person Experiment in Social Influence and Political Mobilization' (2012) 489 Nature 295.

[229] Mark Andrejevic and Kelly Gates, 'Big Data Surveillance: Introduction' (2014) 12 Surveillance & Society 185, 190.

[230] For example, small business have limited access to many valuable databases. Bart Custers and Daniel Bachlechner, 'Advancing the EU Data Economy: Conditions for Realizing the Full of Potential of Data Reuse' (forthcoming in 2018) *Information Policy* 10-11.

[231] Monique Goyens, director general of the European Consumer Organisation, Welcome speech at the EDPS-BEUC conference (Brussels, 29 September 2016) <https://edps.europa.eu/data-protection/our-work/publications/events/edps-beuc-conference-big-data-individual-rights-and_de> accessed 28 May 2018.

[232] In March 2016, the Norwegian Consumer Council filed a complaint with the Norwegian regarding unfair contractual terms in the Terms of Use for the mobile application Tinder. As a result, Tinder later amended the disputable parts of the terms. David Meyer, 'Tinder Is in Trouble Over Its "Unfair" User Terms' *Fortune* (3 March 2016) <http://fortune.com/2016/03/03/tinder-norway-trouble/> accessed 28 May 2018. For a more detailed analysis of Tinder and other apps' terms by the Norwegian Consumer Council see Forbrukerrådet, 'Appfail? Threats to Consumers in Mobile Apps' (2016).

[233] Aleks, Jakulin (@aleksj), 'Why let Google show excerpts of your content in their search results without partaking in the lucrative search advertising revenue?' *Twitter* (April 29, 2016) <https://twitter.com/aleksj/status/725998664687206400> accessed 26 May 2018.

The asymmetry becomes even more apparent when personal data is processed as part of decision-making. Data controllers are able to leverage the collected personal data when they make commercial decisions, whereas individuals have little overview of the process. For instance, based on a personal data analysis, employers are able to determine employees' performance scores. As a consequence, individuals may face a lower salary or a risk of being fired.[234] Because such decisions are typically made on a multi-factor and multilevel analysis of workers' data, an individual may have trouble identifying what exactly is included in this performance that leads to such a 'verdict'.[235]

2.4.2.5. Discrimination

The key objective of *data-driven decision-making* is to differentiate between individuals. Clearly, such decisions can have important consequences for individuals and can work to both their advantage and disadvantage. Certain practices are legally allowed, though it could be argued that they are ethically disputable. For example, some online platforms are able to use the information collected by consumers to their disadvantage: by setting the price as close as possible to the maximum price that someone is willing to pay, they are able to exploit consumers' price sensitivity.[236] This is an example of price discrimination, which may become increasingly aggressive given the level of dataveillance on the Internet.[237]

However, data-driven decisions can also lead to *discriminatory practices* that cross the boundaries of what is legally acceptable. Discrimination that occurs when people are treated differently on the basis of protected grounds is prohibited regardless of whether it happens in a direct or indirect way.[238] An employer may refuse a candidate because an Internet (social media) search reveals how old she is. She may be in her 60s, and therefore too close to retirement, or she may be in her 30s, and therefore too likely to become pregnant. This would constitute illegal discrimination on the grounds of age or sex.[239] Data-driven decision-making may also lead to hidden discrimination. Group profiles inferred from big data are often used as a tool to make decisions about the members of the group, but not every group characteristic can justify different treatment. Characteristics such as address code can be legitimate factors according to which to differentiate, but they might mask ethnicity or religion – both of which are protected grounds.[240]

2.5. Conclusions

Chapter 2 answered the first research sub-question regarding the rise of the data-driven economy and the consequences for individuals. The chapter demonstrated that the world has entered a new era. Data-driven technologies and data analytics have spread across the economy, penetrating even some of the most traditional industries. In this growing data economy, personal data is treated as a highly valuable source, giving the data-driven firms a competitive edge. Individuals too have profited from

---

[234] Sandra Wachter, Brent Mittelstadt and Luciano Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law.
[235] Custers and Ursic, 'Worker Privacy in a Digitalized World under European Law' 340.
[236] Bernasek and Mongan (2015).
[237] Price discrimination and price differentiation are synonyms in economic jargon.
[238] Francesca Bosco and others, 'Profiling Technologies and Fundamental Rights and Values: Regulatory Challenges and Perspectives from European Data Protection Authorities' in Serge Gutwirth, Ronald Leenes and Paul de Hert (eds), *Reforming European Data Protection Law* (2015) 19.
[239] Lynskey (2015) 199.
[240] Custers (2004) 114.

the developments in technology: the prices of certain products have dropped; ordinary consumers have gained access to some sophisticated innovations; and the use of data has improved their lives. However, the intense personal data use has had some negative consequences such as privacy violations, information asymmetries, and imbalance of market power. As these harmful effects have not yet been fully explored, it is important that they are balanced carefully together with the positive outcomes of the data economy. Indeed, the profitable nature of data processing always has a reverse side, namely the need for better protection of personal data. This puts some limitations on what the industry can and cannot do, and law has an important role in setting the right boundaries. The following chapter provides an overview of some key legal rules that apply to the processing of personal data. Subsequently, the thesis is narrowed down to legal provisions that are addressed directly to individuals and concerned with (the extent of) their control over personal data.