D4.1
Results of the gap analysis

Ethical and Societal Implications of Data Sciences

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731873

e-SIDES – Ethical and Societal Implications of Data Sciences

Data-driven innovation is deeply transforming society and the economy. Although there are potentially enormous economic and social benefits this innovation also brings new challenges for individual and collective privacy, security, as well as democracy and participation. The main objective of the CSA e-SIDES is to complement the research on privacy-preserving big data technologies, by analyzing, mapping and clearly identifying the main societal and ethical challenges emerging from the adoption of big data technologies, conforming to the principles of responsible research and innovation; setting up and organizing a sustainable dialogue between industry, research and social actors, as well as networking with the main Research and Innovation Actions and Large Scale Pilots and other framework program projects interested in these issues. It will investigate stakeholders' concerns, and collect their input, framing these results in a clear conceptual framework showing the potential trade-offs between conflicting needs and providing a basis to validate privacy-preserving technologies. It will prepare and widely disseminate community shared conclusions and recommendations highlighting the best way to ultimately build confidence of citizens and businesses towards big data and the data economy.

D4.1 Results of the gap analysis

Work package	WP 4 – Design requirements for new technologies
Lead authors	Karolina La Fors (Leiden University) Alan M. Sears (Leiden University)
Contributing authors	Daniel Bachlechner (Fraunhofer ISI) Michael Friedewald (Fraunhofer ISI) Jana Weitkamp (Fraunhofer ISI) Melek Akca Prill (Fraunhofer ISI) Bart Custers (Leiden University)
Internal review	Richard Stevens (IDC)
Due Date	M21 (September 2018)
Date	15 Oct 2018
Version	39 (final)
Type	Report
Dissemination level	Public

This document is Deliverable 4.1 of Work Package 4 of the e-SIDES project on Ethical and Societal Implications of Data Science. e-SIDES is an EU funded Coordination and Support Action (CSA) that complements Research and Innovation Actions (RIAs) on privacy-preserving big data technologies by exploring the societal and ethical implications of big data technologies and providing a broad basis and wider context to validate privacy-preserving technologies. All interested stakeholders are invited to visit www.e-sides.eu for further information about the e-SIDES results and initiatives.

Executive Summary

This Deliverable provides a gap analysis related to the findings of Deliverable 2.2 and Deliverable 3.2 of the e-SIDES project. Deliverable 2.2 provides a general assessment of ethical, legal, societal and economic issues that emerge in different big data contexts in general, while Deliverable 3.2 provides a technology-specific assessment of classes of currently existing privacy-preserving technologies including their effectiveness in addressing ethical and societal issues and the challenges that arise in their implementation. In Deliverable 3.2 we concluded that ethical and societal issues remain present, mainly because available technologies are not widely integrated into big data solutions. This implementation gap is the focus of this report.

More generally, two broad categories of gaps can be distinguished. First, there may be issues for which no effective solutions exist. Second, there are issues for which solutions do exist, but those solutions are not used or implemented in many cases. The first category could be addressed by further developing privacy-preserving technologies and implementing big data solutions that use them. For the second category, there is less of a need for technological development; rather there is more of a need for finding and addressing the reasons for which the existing and available technologies are not used and implemented in big data solutions. With insight into the ethical, legal, societal and economic reasons behind this, next steps can be formulated to address the reasons and attempt to close the implementation gap.

Consequently, in this deliverable we highlight the ethical, legal, societal and economic aspects encompassing the reasons for the implementation gap. In our identification of reasons behind the gap, we were guided by controversies around the sensitivity of data, the rapid evolution of technologies, the possibility of unforeseen implications and the exceptions to the general rules.

A total of four legal and ethical aspects are distinguished: Privacy by design, Sensitive data, Inferred data, and Liability and ethical responsibility. With regard to societal and economic aspects, the following six reasons are discussed: Costs and benefits, Business models, Public attention, Economic value, Cultural fit, and Skill level. Additionally, we conducted an analysis of the media coverage in the period from shortly before the introduction until a few of weeks after of the introduction of the GDPR with a focus on the United Kingdom and Germany. In this analysis, we paid particular attention to possible explanations for the implementation gap.

With respect to reasons for the limited introduction of privacy-preserving technologies from the legal and ethical aspects we found that:

- The flexible interpretation of privacy and privacy-preserving technologies is both a blessing and a curse for practitioners.
- Specific rules for the protection of special categories of data are embraced to a different extent by professionals. Some healthcare professionals, for instance, see strict privacy preservation as an impediment for epidemiological research.
- In terms of inferred data, the GDPR has limitations and legal gaps. However, we found that for certain cases of inferred data, the new rights of the GDPR can offer remedy. By using a broad interpretation of the right to data portability, for instance, challenges arising from the buying and reselling of EU residents' data by companies in third countries could be addressed.

- In terms of legal liability and ethical responsibility, the entire big data value chain shows gaps, whereas it should embrace privacy preservation as a shared responsibility. Beyond the legal liability of stakeholders, the levelling of responsibilities should correspond with the strength of a given party involved in the data value chain. This degree of liability and responsibility should be made mutually transparent for all the stakeholders involved.
- Regional differences in data protection and antitrust or competition law regimes between the EU and the US also influence the extent to which privacy-preserving technologies are embraced. Harmonisation in favour of a more consumer and data subject friendly approach would be welcome also to balance the market power of large big data companies. The legal regimes of data protection, which should help to stem data breaches, and competition law, which should prevent larger companies from utilising a dominant position in the market, could complement and mutually strengthen each other in this respect.

With respect to reasons for the limited introduction of privacy-preserving technologies from economic and social aspects we found that:

- Adding privacy-preserving technologies to big data solutions leads to additional costs for solution developers and users. These costs must be offset by the expected benefits. Only then can the integration of technologies make sense from an economic point of view. Examples for costs that may be affected include costs for hardware and software as well as costs related to user inconvenience. Benefits may include a reduced need for staff to deal with privacy breaches (e.g., when settling disputes) and an improved reputation.
- Putting privacy principles such as purpose limitation or data minimisation into practice may be in conflict with current or desired business models. Moreover, conflicts may arise from different treatments of special categories of data and legal rules governing decision-making processes. Closely related to business model conflicts is the trade-off between privacy protection and the utility of data. It was found that increased data protection limits flexibility and innovation in certain contexts.
- Taking privacy and transparency seriously, and making this public has the potential to allow for competitive differentiation. The term ‘privacy as a strategy’ has been used to refer to the phenomenon of using privacy preservation approaches for competitive differentiation. Apple is a well-known supporter of this strategy, but there are many others. Transparency is particularly as issue in settings where there are hardly any alternatives for specific services available.
- Understanding the value of privacy for individuals is essential as this allows designing appropriate big data solutions. Individuals have the potential to exert significant pressure on actors in the data value chain. The literature shows that privacy concerns and expectations are remarkably context-dependent and very difficult to predict. However, where there is a privacy difference between companies, the privacy-unfriendly company typically obtains a greater market share.
- Acknowledging that privacy preferences and practices vary among nations and regions is important. A universal regulatory approach to information privacy would ignore cultural and societal differences. A key question is whether big data is considered to rather lead to de-individualisation and discrimination, or to personalisation. It was found that cultural values have an impact on the extent to which errors in databases and unauthorised secondary use raise privacy concerns.

- Adapting to a new mind-set seems to be necessary as data has become a strategic business asset and privacy a threatened value. Today, big data professionals must have skills ranging from math and statistics, machine learning, decision management and computer science to data ethics, law and information security. Several of these skills are essential for developers to make sure privacy-preserving features are properly integrated into big data solutions.

As a conclusion of our analysis of ethical legal, economic and societal reasons leading up to and stemming from the limited or scarce introduction of privacy-preserving technologies, we also stress the need for the constant reassessment of privacy-preserving technologies with respect to their design and use in relation to the social, ethical, legal and economic effects on persons and society at large. As we have demonstrated, these technologies alone cannot address all ethical and societal issues and the challenges that arise from the very implementation of big data solutions. But legal, ethical, economic and societal changes are needed in order to facilitate transformations towards a more privacy-preserving data exchange and usage that capitalizes on the value of big data but respects the ethical, legal and societal limitations of exploiting that data.

Table of Contents

Figures

Tables

Abbreviations

CJEU	Court of Justice of the European Union
DPD	Data Protection Directive
EFF	Electronic Frontier Foundation
ENISA	European Union Agency for Network and Information Security
FTC	Federal Trade Commission
GDPR	General Data Protection Regulation
OECD	Organisation for Economic Co-operation and Development
SMEs	small and medium-sized enterprises
TFEU	Treaty of the European Union
WP29	Article 29 Working Party

1. Introduction

This document is Deliverable 4.1 of Work Package 4 of the e-SIDES project on the Ethical and Societal Implications of Data Science. e-SIDES is an EU funded Coordination and Support Action (CSA) that complements Research and Innovation Actions (RIAs) on privacy-preserving big data technologies by exploring the societal and ethical implications of big data technologies and providing a broad basis and wider context to validate privacy-preserving technologies. The current deliverable provides a gap analysis between the findings of Deliverable 2.2 and Deliverable 3.2 of E-SIDES. Deliverable 2.2 had the objective to provide a general assessment of ethical, legal, societal and economic issues that emerge in different big data contexts in general. Deliverable 3.2 provided a technology-specific assessment of classes of currently existing privacy-preserving big data solutions including their effectiveness in addressing ethical and societal issues and the challenges that arise from the very implementation of these solutions. In Deliverable 3.2 we concluded that despite many existing solutions, ethical and societal issues remain present. This gap between issues and solutions is the focus of this report.

Two types of gaps can be distinguished. First, there may be issues for which there are no effective solutions. Second, there may be issues for which solutions do exist, but those solutions are not used or implemented. The first type of gap could be addressed by further developing design requirements for privacy preserving big data technologies and innovating new types of solutions. For the second type of gap, further technological developments are not required, but rather finding the reasons for which the existing and available solutions are not used and implemented is needed. Perhaps the solutions are unknown, too expensive or not considered a priority. Depending on the reasons behind this, next steps can be formulated to address the ethical, legal, societal and economic issues.

In this report, we will focus mostly on the second type of gap, as the assessment of existing privacy-preserving technologies conducted within the scope of Work Package 3 suggests that there are quite comprehensive privacy-preserving technologies available but ethically, legally, socially and economically motivated reasons exist that hamper their implementation. Privacy by design is regularly mentioned in related debates as a key approach to systems engineering, but it does not seem to be used comprehensively by developers of big data solutions. At the same time, however, companies, including developers of big data solutions, increasingly seem to try to brand themselves as privacy protectors.[1] Still, existing privacy-preserving technologies, such as homomorphic encryption, blockchains, anonymisation and others are often not implemented in practice. In certain cases, the needs of industrial and other big data applications are not currently met by these existing technologies[2] and therefore need to be adapted to new contexts.

The assessment of existing technologies conducted within the scope of WP3 suggests that privacy-preserving technologies are integrated only to a limited extent in today's big data solutions. This raises the question why this is the case. Currently, there are limited explanations that show why this is the case or what could be done to increase the practical use of privacy-preserving technologies. This report lists ethical, social, economic and social reasons leading up to the limited introduction of these technologies.

[1] Deliverable 3.2.
[2] For more about this please see Deliverable 3.2.

Regional differences between the EU and the US from a social and economic perspective and concerning the approaches toward privacy-preserving technologies - as we will demonstrate - also tie into economic and social reasons for their limited implementation. Societal reasons for such differences are grounded upon differing organisational traditions (as we also saw above), management cultures, communication lines and others, which are also closely interlinked with the clear differences between EU and US legal regimes regarding privacy and data protection. Although the latter issues demonstrate rather economic and societal perspectives as they relate to corporate culture, organisational, management difficulties. These reasons are also intertwined with legal and ethical aspects.

In line with this the structure of this deliverable looks as follows: In section 2 we will introduce legal and ethical gaps that encompass reflections about the sensitivity of data, the rapid evolution of technologies, the possibility of unforeseen implications and the exceptions to the general rules. In line with this under the section on legal and ethical gaps, four sub-sections are distinguished: Privacy by design, Sensitive data, Inferred data, and Liability and ethical responsibility. In section 3 we highlight the societal and economic gaps and have the following sub-sections: Costs and benefits, Business models, Public attention, Economic value, Cultural fit, and Skill level. In section 4, we provide insights into the limits of technical approaches and in section 5 a media analysis with respect to the time shortly before the introduction and a couple of weeks after of the introduction of the GDPR. In section 6 we provide our conclusions.

2. Legal and ethical aspects

Reasons from ethical and legal factors are also influential regarding the extent to which privacy-preserving technologies are introduced. This chapter examines the following legal and ethical aspects in which reasons for the implementation gap can be found: privacy by design (section 2.1), sensitive data (section 2.2), inferred data (section 2.3) and liability and ethical responsibility (section 2.4).

2.1. Privacy by design

The most obvious starting point from a legal perspective is the GDPR provision (art. 25) on Privacy by design, which is closely related to the concept of privacy-preserving technologies as it requires privacy safeguards to be integrated in technological solutions. Technologies to preserve or enhance privacy were first discussed at length in a 1995 report that resulted from a joint project set up by the Dutch Data Protection Authority and the Ontario Information Commissioner[3]. The seven foundational principles of privacy by design[4] developed in 2011 are highly relevant in the context of big data solutions and can be put into practice by integrating privacy-preserving technologies:

- *Proactive* not Reactive; *Preventative* not Remedial: Interviewees who participated in the assessment of existing privacy-preserving technologies clearly stated that technologies, in contrast to many organisational measures, tend to be proactive and preventive. The technologies aim to prevent privacy-invasive events from happening.
- Privacy as the *Default*: Interviewees highlighted when talking about privacy in the context of big data that the strongest party should have the biggest responsibilities. It was criticised that responsibility is to some extent pushed to the individual who may not fully understand what is happening. Privacy-preserving technologies contribute to the automatic protection of personal data.
- Privacy *Embedded* into Design: It was clearly stated by the interviewees that privacy preservation is unlikely to work as an add-on. It must be an essential component of the core functionality of a big data solution. The ignorance or circumvention of technologies must be difficult if not impossible. Related aspects are discussed in more detail in the section on customers and users (see section 2).
- *Full* Functionality - Positive Sum, not Zero-Sum: It is important that unnecessary trade-offs are avoided. The assessment of existing privacy-preserving technologies clearly shows that technologies such as multi-party computation or homomorphic encryption allow for both the protection of privacy and the utility of data. Moreover, it was clearly stated by interviewees that technologies focusing on access, portability or user control tend to be beneficial for both the users of the big data solutions as well as the data subjects.
- End-to-End Lifecycle Protection: The entire privacy-preservation process is only as good as the weakest link in the chain. At the same time, the distributed computing common in the era of big

[3] "Privacy-Enhancing Technologies: The Path to Anonymity", http://www.ontla.on.ca/library/repository/mon/10000/184530.pdf

[4] "Privacy by Design in Law, Policy and Practice", http://www.ontla.on.ca/library/repository/mon/25008/312239.pdf

data opens up additional opportunities for data breaches. Some of the technologies assessed, particularly those related to encryption, access control and policy enforcement, help ensuring end-to-end lifecycle protection.

- Visibility and Transparency: Visibility and transparency are important with respect to both the technologies and organisational measures used to preserve privacy, and the big data solutions in general. Technologies related to transparency and data provenance, which are among the technologies assessed in WP3, clearly contribute to improving the visibility and transparency of big data solutions.
- Respect for User Privacy: The interests of individuals must be held in the highest regard when developing big data solutions that are in line with the privacy by design approach. In combination with organisational measures, the privacy-preserving technologies that were assessed within the scope of the project can make a significant contribution. What is essential, however, is that they are integrated into big data solutions and used.

Although these principles create still the bedrock for privacy by design innovation, critics argue that their flexible interpretation is both a blessing and a curse for practitioners.[5]

Hustinx[6] states that the need for privacy by design could never be better illustrated than by the increasing number of data security breaches. The fact that there are still high numbers of breaches that could have been prevented by reasonable and rather low-cost measures indicates that more efforts have to be made (see section 2.4 for more details about data breaches and their effect on the use of privacy-preserving technologies).

The new EU data protection regime is since May 25th 2018 in force. Its privacy by design and by default provision explicitly imposes a legal obligation onto data processing organisations that process data of EU residents to "implement appropriate technical and organisational measures [...] which are designed to implement data-protection principles [...] in an effective manner and to integrate the necessary safeguards" (Article 25). The GDPR includes the principle of privacy by default. However, the Directive 95/46/EC (Article 17) already required “appropriate technical and organisational measures to protect personal data against accidental or unlawful [...] forms of processing” (see section 2.4 for more details on the relevance of legislation on the use of privacy-preserving technologies and the section on media analysis for a discussion of specific effects of the GDPR).

According to Gürses & Troncoso[7], the concept of privacy by design has received considerable attention in policy circles but the actual design, implementation and integration remains an open question. Different parties have proposed privacy by design methodologies[8]. These methodologies, however, address the

[5] Strategic Privacy by design, an interview with Jason Cronk - < https://teachprivacy.com/strategic-privacy-by-design/>

[6] Peter Hustinx, "Privacy by design: delivering the promises", doi:10.1007/s12394-010-0061-z

[7] Seda Gürses, Carmela Troncoso, Claudia Diaz, "Engineering Privacy by Design Reloaded"

[8] Proposals were made by Thibaud Anignac, Daniel Le Métayer, "Privacy by Design: From Technologies to Architectures", doi:10.1007/978-3-319-06749-0_1 and Jaap-Henk Hoepman, "Privacy Design Strategies", doi:10.1007/978-3-642-55415-5_38

engineering aspects of privacy by design by pointing to design strategies, without explaining how these strategies should be applied to build privacy-preserving solutions.

In an ENISA report, D'Acquisto et al.[9] point out that the concept of privacy by design is essential to identify privacy requirements early and subsequently implement the necessary technical and organizational measures. The report acknowledges that putting the privacy by design principles into practice in the big data context is not straightforward. The authors state that further guidance is required especially when many players are involved and privacy can be compromised at various points of the big data value chain. Moreover, they see data protection authorities as well as the developers and users of big data solutions having a shared responsibility in defining how privacy by design can be practically implemented.

In accordance with the legal frameworks referenced above, D'Acquisto et al. stress that privacy by design cannot be reduced to a collection of general principles or the implementation of technologies that preserve privacy but that it must be seen as a process that involves technological and organisational components. Waldman[10] equally advocates for the 'institutionalization of privacy' that involves legal, technological, organizational and individual stakeholders.

Privacy by design seems to be in conflict with key aspects of big data. Data reuse, for instance, goes against the principle of purpose limitation, the massive collection of data opposes data minimisation, and the involvement of many controllers and the complicated interaction between them makes transparency and control difficult. These may be key barriers for the implementation of privacy by design in different contexts of big data. Whereas D'Acquisto et al. argue that privacy and big data can go well together if privacy is considered a core value of big data, others are less confident in that regard. Tene & Polonetsky[11] and the US President's Council of Advisors on Science and Technology (PCAST),[12] for instance, stress that the opportunities of reidentification have undermined the faith placed in anonymisation technologies. The implications can be significant as anonymisation has become a key component of many business models, most notably in the context of health care and online behavioural advertising, web-surfing and others (see section **Error! Reference source not found.** for more details on business model conflicts).

To conclude, the new regulations imposing privacy by design and privacy by default obligations on data processing companies in Europe and the current privacy regulations in the US follow, what Schwartz & Solove outline as the 'notice and choice' framework.[13] This is a discrepancy, because the notice and choice framework means that through the privacy protection rules a large burden of responsibility is still put back upon the data subject in managing his/her own privacy. Individuals are pressed to take action in defence

[9] D'Acquisto et al., "Privacy by design in big data", ENISA.

[10] Waldman, Ari Ezra, Designing Without Privacy (March 31, 2017). Houston Law Review, Vol. 55, No. 659, 2018; NYLS Legal Studies Research Paper No. 2944185; 55 Houston L. Rev. 659 (2018). Available at SSRN: https://ssrn.com/abstract=2944185.

[11] Omer Tene, Jules Polonetsky, "Privacy in the Age of Big Data: A Time for Big Decisions".

[12] PCAST, "Big Data and Privacy: A Technological Perspective", https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf

[13] Paul M. Schwartz, Daniel J. Solove, "Notice and Choice: Implications for Digital Marketing to Youth", Memo prepared for The Second NPLAN/BMSG Meeting on Digital Media and Marketing to Children, https://www.changelabsolutions.org/sites/default/files/documents/Notice_and_choice.pdf; Paul M. Schwartz, Daniel J. Solove, "The PII Problem: Privacy and a New Concept of Personally Identifiable Information", New York University Law Review 86.

of their privacy based on notifications about the potential whereabouts of data and its potential (re)use. So, the framework "envisages a consumer who self-manages privacy by dealing strategically with data collection entities", but in real life the growing amount of studies show that in the big data era where personal data reuse became ordinary practice "users' actual knowledge about the mechanisms behind the new digital economy does not sufficiently equip them with the tools to protect their privacy".

2.2. Sensitive data

The strictest legal regime in the GDPR applies to sensitive data, which is in line with the increased risks for issues and challenges related to sensitive data. Therefore, in cases of processing sensitive data, extensive use of sophisticated technological solutions might be expected. The GDPR brought along significant changes also with respect to the protection of sensitive data. Art. 9 of the GDPR sets out the following personal data to be subject to specific processing conditions (as being sensitive):

- "*personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;*
- *trade-union membership;*
- *genetic data, biometric data processed solely to identify a human being;*
- *health-related data;*
- *data concerning a person's sex life or sexual orientation."*

This list covers an extensive set of data that in case it were used as inputted data, could already reveal information on sensitive personal traits. Based on such traits judgements can be made which can jeopardize human dignity and cause prejudice based on religious or ethnic background or sexual orientation. Therefore Art. 9 is an important step forward in itself within the GDPR. However not all professionals and disciplines welcomed to the same extent the form of sensitive data protection this new GDPR article provided. For instance, epidemiologists raised their concerns that the current form of protection provided by the GDPR could limit the efficacy of epidemiology research in the future (Nyrén, Stenbeck, & Grönberg, 2014). The need for a strong protection of privacy and identity of patients is clear as the medical profession for long had aimed to respect human welfare in general and values such as benevolence, non-maleficence, justice and dignity enshrined by bioethics and the institution of the Hippocratic Oath more specifically. Therefore, to design privacy-preserving technologies that address such ethical and legal aspects had been a crucial need. Such projects, as MyHealthMyData[14] for instance, aimed to develop blockchain technology that is aimed at allowing patients to self-govern their patient identity by offering dynamic consent management for patients. Although Carechain,[15] as the specific blockchain system is called, is currently under development it is aimed at addressing different legal issues when it comes to the implementation of these specific (blockchain-based) privacy-preserving technologies, because it will make data transfer traceable and the right to data portability something to easily effectuate. On the other hand, legal and ethical issues may remain for individuals in terms of how to

[14] My Health My Data,- <http://www.myhealthmydata.eu/>

[15] HIMSS Europe, "A Nordic way to blockchain in healthcare", 26 February 2018, https://www.himss.eu/himss-blog/nordic-way-blockchain-healthcare.

strengthen their control when it comes to the automated nature of smart contracts or how to realize their right to be forgotten in a limited but mutually transparent block-chain structure.[16] When it comes to healthcare data and data that falls within the scope of the vital interest of the data subject such as biometric data or data about sexual orientation, Recital 46 allows for the processing of such data when it serves the "vital interest of the data subject".[17]

Yet, the vital interest of the data subject is more challenging to demonstrate, when it comes to the religious, racial and ethnic background or political orientation of a person. Still similarly to healthcare data for such pieces of data, data protection rules can be easier complied with when data input reveals information on such personal traits and not as a consequence of data analytics sensitive personal traits become revealed. When as the consequences of data analytics sensitive personal traits listed under Art. 9 of the GDPR become revealed the effectivity of the GDPR becomes weaker. As we will discuss this in the following section, inferred data under GDPR is a muddy field and especially complicated when it comes to inferred data that reveals sensitive information about a person. Cases of commercial profiling of customers demonstrate that even when privacy-preserving technologies, such as encryptions, are in place ethical values, such as human dignity can be damaged because specific purchasing information of a person after analytics can turn out to be inferred data that reveals sensitive traits about a person for others. The case of a pregnant woman, who wanted to keep her early pregnancy as a secret in front her parents was exposed to commercial profiling advertisements via regular mail, the content of the advertisement however ultimately revealed her pregnant state for her parents without her intention and beyond her control.[18] Such instances demonstrate remaining ethical and legal aspects or challenges even after certain privacy-preserving technologies are in place already.

2.3. Inferred data

Another aspect to be considered is that of inferred data, in which current EU regulation lags behind with respect to providing clear legal remedy for the unforeseen implications of data analysis that is based upon data that was gathered for different purposes and reused for previously undefined purposes. Inferred data is all data that stems from big data analysis where input data comes from a large variety of sources and can go through multiple analytic processes until data points become assembled and analysed for a new purpose. From a personal data protection perspective inferred data is highly important because for persons in the big data era it becomes increasingly impossible not to become involved into big data analytic decisions. The vast amount of data sources and their linkability points towards a direction when

[16] Ibid.

[17] Recitals are the part of the act which contains the statement of reasons for its adoption; they help to provide context for the provision and are nonbinding.

[18] Hill, K. "How Target figured out a girl was pregnant before her father did", Forbes (16 February 2012), https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#175a21896668.

everybody would be identifiable through various data relations and as Purtova refers to personal data protection law would apply to everything[19].

Therefore, and especially with respect to inferred data, where the likeliness for the identifiability of persons only increases through data reuse, the GDPR has loopholes. As Zarsky states the GDPR is "incompatible with the data environment that the availability of Big Data generates"[20]. Although the GDPR puts data minimization and risk prevention requirements upfront for companies, companies can circumvent legal compliance needs. Robert Medge an expert of the MyData platform refers to five major loopholes of the GDPR companies might be interested in and we will zoom here into three from the perspective of inferred data: the data controller's location, data beyond the EU and legitimate interest. The MyData platform[21] promotes a high-level approach for governing personal data in human centric way that is based on legal rights and principles but goes beyond them by embracing ethical values. There is a fundamental contradiction between big data analytics and restricting the inferred and unintended use of data.

1) The data controller's location can be outside the EU

It remains unclear whether organisations "offering goods and services" will be included in the scope of Art. 3(2)a. In case a company, for instance, is located within the EU and offers goods and/or services and only the catalogue of those goods and services could customers consult. Yet, to access the catalogue a link on each service would direct customers to a third-party outside the EU. This practically means that during the actual personal data processing no GDPR compliance requirement would apply, since the third-party which handles the data resides beyond the EU. But whether or not offering goods and services either from within or outside the EU would fall under the GDPR needs to be decide on a case by case basis, as Pammer v Reederei Karl Schlüter GmbH & Co and Hotel Alpenhof GesmbH v Heller (Joined cases C-585/08 and C-144/09) of the Court of Justice of the European Union (CJEU) also demonstrate.[22]

[19] Purtova, N. (2018) The law of everything. Broad concept of personal data and future of EU data protection law, Law, Innovation and Technology, 10:1, 40-81,

[20] Zarsky, T. (2017) Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review*, Vol. 47, No. 4(2).

[21] Poikola, A. Kuikkaniemi, K., Honko, H. (2018) MyData – A Nordic Model for human-centered personal data management and processing - Open Knowledge Finland's MyData Group, Finnish Ministry of Transport and Communication -<https://www.lvm.fi/documents/20181/859937/MyData-nordic-model/2e9b4eb0-68d7-463b-9460-821493449a63?version=1.0>

[22] Joined Cases C-585/08 and C-144/09 of the CJEU, OJ C 44, 21.2.2009.

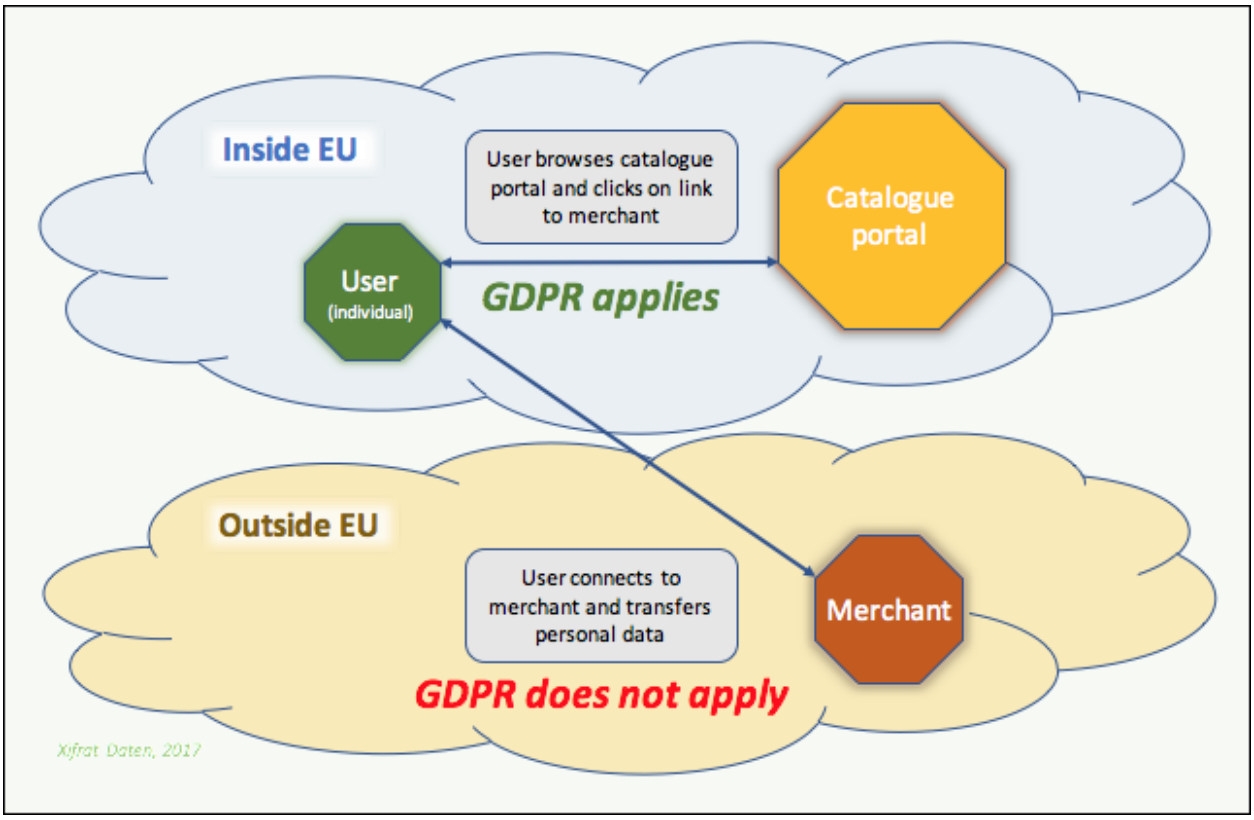

Figure 1: Applicability of the GDPR[23]

2) Data beyond the EU

With respect to data offered to EU citizens, if a company beyond the EU offers goods or services for EU citizens ought to comply with the GDPR and should consequently ask for the consent of customers as Art. 3(2) prescribes that the data processing "relates to the offering of services and goods". When this third country company, however, sells EU customers data to another company in a third country and that company would not use the data for 'offering services and goods to EU citizens', than the data sold to this second company would fall beyond the scope of GDPR as well as the purchased data with it.

An interesting possibility to go after such company is offered perhaps by the Article 29 Working Party's guideline on the Right to Data Portability. Art. 20 of the GDPR sets out the right to data portability. This right *"allows for data subjects to receive the personal data that they have provided to a data controller, in a structured, commonly used and machine-readable format, and to transmit those data to another data controller without hindrance"*[24]. However, the Art. 29 Working Party in its opinion embraces a broader scope of data that should fall under the right to data portability. According to the Article 29 Working Party not only data 'provided to a data controller' but also data observed by the data controller would also fall

[23] Image is from <https://medium.com/mydata/five-loopholes-in-the-gdpr-367443c4248b>.

[24] Article 29 Working Party Opinion on The right to data portability - <http://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611233>

in the scope of this right. Critics of this opinion of the WP29 stress that such a broad interpretation would create for data processing companies too severe of a burden[25]. They would grapple with defining what precautionary measures to choose in light of what counts as 'observed data' and what not. Such a broad interpretation of the WP29 on the right to data portability could address, however, challenges arising from the use of inferred data and specifically from the reselling and buying of EU residents' data by companies in third countries.

3) Legitimate interest

Article 6.1 of the GDPR defines the lawful grounds for data processing as follows:

- **Consent** of the data subject
- Processing is necessary for the **performance of a contract** with the data subject or to take steps to enter into a contract
- Processing is necessary for compliance with a **legal obligation**
- Processing is necessary to protect the **vital interests** of a data subject or another person
- Processing is necessary for the performance of a **task carried out in the public interest** or in the **exercise of official authority** vested in the controller
- Necessary for the purposes of **legitimate interests pursued by the controller or a third party**, except where such interests are overridden by the interests, rights or freedoms of the data subject. This latter condition does not apply for processing carried out by government authorities in order to fulfill their tasks and duties. Such conditions, therefore, do not apply for law enforcement authorities when they process data for the purposes of criminal investigation or prevention. For such activities the EU's Police and Criminal Justice Data Protection Directive[26] applies.

Madge observes that the legitimate interest tests along the abovementioned grounds are difficult to conduct, because it is impossible to define precise rules for conducting "a balance of interests assessment, combined with a procedure that theoretically puts the burden of proof on the controller but in practice leaves controllers almost unsupervised."[27] Therefore, although the above conditions apply, it becomes a matter of case-by-case assessment whether data processed for one purpose was indeed legitimate. Yet, in the case of inferred data (which we regard as data that is processed for renewed purposes) to conduct a legitimate interest test as well as a balance of interest assessment looks very troublesome. Probably the most valuable help for assessing the implications and legitimate interest of the processing of inferred data grants the GDPR's through the privacy impact assessment prescription.

The GDPR acknowledges that data protection rights are not absolute rights but must also be balanced with such fundamental rights as the freedom to conduct business. Member states can introduce

[25] Meyer, D. (2017) European Commission, experts uneasy over WP29 data portability interpretation - <https://iapp.org/news/a/european-commission-experts-uneasy-over-wp29-data-portability-interpretation-1/>
[26] Directive (EU) 2016/680 of the European Parliament and of the Council 27 April 2016 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32016L0680>
[27] Robert Madge, "Five loopholes in the GDPR", Medium (27 August 2017) <https://medium.com/mydata/five-loopholes-in-the-gdpr-367443c4248b>.

derogations and exceptions under the GDPR's derogations and special conditions. The latter can result in different member states with stronger and less strong data protection rules. Web browsing is a big data context that exemplifies the possibility of unforeseen implications of data processing and in which there may even be sensitive personal data that is transferred. This context is most apparent in the collection and use of consumer data. This commonly occurs through the use of third-party web tracking, in which 3rd party cookies are used to monitor user behaviour across sites, usually with the intention to present targeted or relevant ads to the user.[28] While this practice originated in the late 1990s, the industry has grown to include large players such as Google in 2009.[29] However, some sites also collect large amounts of data in relation to a user's behaviour on their own site. For instance, Facebook collects data on a user's mouse or cursor movements.[30]

Despite the positive new developments brought along by the GDPR and the holes in the GDPR's regime with respect to inferred data are undeniable. Yet, a large portion of these holes can be addressed by how courts will interpret the rights offered by the GDPR and the capacities of data subjects in demonstrating harm because of data inferred about them.

2.4. Liability and ethical responsibility

In general, organisations collecting, using and distributing data are responsible for data management and tasks related to privacy preservation such as anonymisation and encryption. Under the GDPR, organisations that are considered data controllers or processors are obliged to "implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk".[31]

In addition to any legal obligation to implement privacy-preserving technologies, data controllers and processors have an ethical responsibility to utilise privacy-preserving technologies so as to address ethical issues that arise in the use of big data technologies, such as autonomy, accountability, trustworthiness, privacy, and dignity, among others.[32] However, even if measures are taken based on the interpretation of applicable laws and an understanding of ethical responsibilities, it does not exclude the possibility of data breaches or other issues that can affect individuals. The impact of such incidents should nevertheless be

[28] V. Toubiana et al., Adnostic: Privacy Preserving Targeted Advertising, Adnostic Whitepaper (Stanford, NYU), https://crypto.stanford.edu/adnostic/adnostic-ndss.pdf.

[29] Google announce in March 2009 that its AdSense service would begin to present ads based on a user's behaviour while browsing. Ibid. See also, Google adsense. https://adsense.blogspot.com/2009/03/driving-monetization-with-ads-that.html

[30] J. Kanter, "Facebook is tracking you in ways you never knew — here's the crazy amount of data it sucks up", Business Insider, <https://www.businessinsider.com/facebook-reveals-all-the-way-it-tracks-user-behaviour-2018-6/?international=true&r=US>.

[31] Art. 32 of the GDPR. These measures may include: "(a) the pseudonymisation and encryption of personal data; (b) the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; (c) the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; (d) a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing."

[32] For a full list of the identified ethical issues in the use of big data technologies, see e-SIDES Deliverable 2.2.

minimised if these considerations are taken into account when deciding which, or the extent to which, privacy-preserving technologies are implemented.

Others with expertise in online privacy have stated that "*consumers need to protect themselves because nobody else will be doing it for them*".[33] Looking at the entire data value chain, it is quite clear that privacy preservation must be a shared responsibility. However, as a technological advisor stated in relation to an interview conducted for Deliverable 3.2, the strongest party should have the biggest responsibilities.[34] This is largely reflected in the GDPR, which allows data protection authorities to impose fines of up to €10 million, or 2% of the worldwide annual turnover of the preceding financial year, whichever is higher, for non-compliance with a number of provisions, and penalties of up to €20 million, or 4% of the worldwide annual turnover of the preceding financial year, whichever is higher, for others.[35]

According to Tankard, with the right precautions in place, organisations have little to fear.[36] Privacy-preserving technologies related, for instance, to encryption, access control and accountability are considered highly relevant for compliance with the GDPR. Nevertheless, Tankard considers the minimisation of the amount of data collected a good practice. While the GDPR largely harmonised the data protection framework in the EU, it remains to be seen the level to which it is enforced by the respective member states. Under the preceding Data Protection Directive (DPD), there was quite a large variation in enforcement.[37]

However, Albrecht states that the GDPR also puts governments outside the EU under pressure to raise their data protection standards.[38] Countries such as Japan are already discussing provisions similar to those of the GDPR, and companies in the UK are doing their best to make sure the GDPR applies to its full extent even after Brexit. A summary of specific effects of the GDPR on the implementation and use of privacy-preserving technologies is provided in section **Error! Reference source not found.** on the media analysis.

There is also little understanding as to who should be responsible for what and how one party can be sure that the others are trustworthy. It is important that data protection is not considered as "somebody else's problem" as this point of view passes the responsibility from one party to another. In the EU, data protection authorities and governments definitely play an important role in this regard.

The analysis of responsibility in the context of privacy protection focuses on all actors along the data value chain from the data subject to the user of the data. Clarity with respect to liability and responsibility is a key prerequisite for the development of reasonable design requirements for the implementation of big

[33] Taylor Armerding, "The 5 worst big data privacy risks (and how to guard against them)", CSO (14 July 2017). Armerding cites Rebecca Herold, CEO of The Privacy Professor, with this statement in his article.
[34] e-SIDES Deliverable 3.2 assessed a wide range of privacy-preserving technologies.
[35] Art. 83, 4. and 5. GDPR.
[36] Colin Tankard, "What the GDPR means for businesses", doi:10.1016/S1353-4858(16)30056-3
[37] Bart Custers, Francien Dechesne, Alan M. Sears, Tommaso Tani, & Simone van der Hof, "A Comparison of Data Protection Legislation and Policies Across the EU", Computer Law & Security Review, Volume 34, Issue 2, April 2018, pp. 234-243, https://www.sciencedirect.com/science/article/pii/S0267364917302856.
[38] Jan Philipp Albrecht, "How the GDPR Will Change the World", doi:10.21552/EDPL/2016/3/4.

data solutions. Insight into current points of view with respect to liability and responsibility help in understanding why existing privacy-preserving big data solutions are implemented as they are.

Raab & Bennett[39] state that the available knowledge about how privacy protection laws and systems work does not easily relate to the circumstances of particular individuals or groups. Moreover, it is very likely that these patterns of risk, protection and perception vary by sector, country and type of person. According to Raab & Bennett, privacy risks and their distribution have not yet enjoyed a widespread, evidenced discourse that might reveal where the areas of agreement and disagreement lie. Further research is considered necessary to suggest strategies for coping more effectively with risks and fears that arise out of information technology and its applications, and to take into account the disparities among social groups and categories in the protection of their personal information. Raab & Bennett conclude that such research could be useful, among other reasons, to show the effect of certain solutions such as privacy-enhancing technologies and market-based initiatives. Apart from that, deeper insight into the distribution of privacy risks and privacy protection may also allow shedding light on the pros and cons of different approaches in the distribution of responsibility.

The following subsections address liability and ethical responsibility in regards to data breaches and competition law in particular. Regional differences, especially in relation to the United States, will also be examined; many big data companies are based there, and the context may help to illuminate differences in the usage of privacy-preserving technologies.

2.4.1. Data breaches

Data breaches have continued to occur despite laws and regulations that require organisations to implement security measures.[40] Public awareness of security breaches has risen due to increased reporting on the matter and may be further expected to rise due to new data breach notification obligations in the GDPR. Prior to the GDPR, there was no uniform legislation regarding data breach notifications in the EU, except for a specific obligation on electronic communications service providers under the ePrivacy Directive.[41] Art. 33 of the GDPR requires that controllers must notify the relevant supervisory authority within 72 hours after having become aware of a personal data breach, unless the breach is unlikely to result in a risk to the rights and freedoms of the data subjects concerned.[42] Failure to comply with this provision can result in a fine of up to €10 million, or 2% of the worldwide annual revenue of the preceding financial year, whichever is higher.

---

[39] C. D. Raab, C. J. Benett, "The Distribution of Privacy Risks: Who Needs Protection?", doi:10.1080/019722498128719

[40] One such example of a regulation is found in Art. 32 of the GDPR, as mentioned in the preceding section.

[41] Directive 2002/58/EC, as amended in 2009.

[42] The notification must also: "(a) describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned and the categories and approximate number of personal data records concerned; (b) communicate the name and contact details of the data protection officer or other contact point where more information can be obtained; (c) describe the likely consequences of the personal data breach; [and] (d) describe the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects." Art. 33, 2.

According to Culnan & Williams,[43] in the majority of cases, organisations had failed to implement even standard security policies and procedures and were slow to detect breaches. In fact, rather than detecting the breach themselves, more than two out of three companies involved in a study conducted by the authors learned they had suffered a breach from a third party.[44]

According to a report of the Ponemon Institute,[45] victims of breaches, particularly if they suffer negative consequences such as identity theft, are likely to lose trust and confidence in the organisation. Obviously, lost trust will likely cause many customers to terminate their relationship with the organisation, especially if they believe its response to and handling of the security breach is unsatisfactory. Notification laws, according to Romanosky et al.,[46] aim at inducing organisations to invest and improve their security measures in order to avoid data breaches. However, the authors state that managers of companies may believe that their probability of suffering a breach is small enough that they may still not fully appreciate the associated penalties. It may also be that they estimate the net direct and indirect costs of breaches to be quite small, compared to the investments necessary to significantly decrease the probability of those breaches. Interestingly, the extensive use of encryption has been shown to be one of the most important factors in decreasing the cost of a data breach.[47] With developments such as the GDPR entering into force, however, ignoring data protection compliance requirements becomes less attractive.

Acquisti et al.[48] show that there is a statistically significant negative impact of data breaches on a company's market value on the announcement day for the breach.[49] However, they also found that the impact is lower than the one documented in literature for security incidents based, for instance, on viruses or denial-of-service attacks. Acquisti et al. state that data breaches may be more confusing than security breaches as their magnitude, implications and nature tend to be more complicated and often not immediately tractable. Moreover, the authors state that it appears that smaller companies and retail companies are affected to a greater degree than other companies. Acquisti et al. explain that consumers' switching costs are lower in the case of retail companies than they are for other companies such as banks. Evidence of a malicious actor deliberately trying to access the data increased the negative effect. The number of data subjects affected seems to be relevant only in relation to very large data breaches.

In some instances, the inadequate use of privacy-preserving technologies may have significant consequences and result in the loss of sensitive personal data, such as health records. Data such as these are particularly important to secure because of the information that may be gleaned from their release, and yet the number of breaches affecting healthcare organisations increased nearly every year since 2010

[43] Mary, J. Culnan, Cynthia Clark Williams, "How Ethics Can Enhance Organizational Privacy: Lessons from the Choicepoint and TJX Data Breaches", MIS Quarterly

[44] Ibid.

[45] Ponemon Institute, "National Survey on Data Security Breach Notification"

[46] Sasha Romanosky, Rahul Telang, Alessandro Acquisti, "Do Data Breach Disclosure Laws Reduce Identity Theft?", doi:10.1002/pam.20567.

[47] Ponemon Institute, "2018 Cost of a Data Breach Study: Global Overview", p. 22, https://public.dhe.ibm.com/common/ssi/ecm/55/en/55017055usen/2018-global-codb-report_06271811_55017055USEN.pdf.

[48] Alessandro Acquisti et al., "Is There a Cost to Privacy Breaches? An Event Study", ICIS 2016.

[49] This study was conducted in the US.

in the US.[50] The importance is highlighted by the fact patient records have been found to be very valuable on the dark web, and often sell for 300 to 400 USD per record, compared with 1 or 2 USD for financial data such as credit card information.[51]

2.4.2. Competition law

Competition law may also be a driving force in the implementation of privacy-preserving technologies. The Treaty of the European Union (TFEU) underpins competition law in the EU. Article 101 prohibits agreements, decisions by associations of undertakings, and concerted practices that distort competition within the European single market. Additionally, the TFEU in Article 102 forbids the abuse of a dominant position within the internal market. These are given force under Council Regulation 1/2003, as businesses that violate these provisions are liable for up to 10% of their worldwide annual turnover.[52]

It should also be noted that other legal provisions in the EU interface with competition and antitrust law.[53] The EU Charter of Fundamental Rights provides that EU policies shall ensure a high level of consumer protection in Article 38, and there is the freedom to conduct business in Article 16, which entails that small and medium-sized enterprises (SMEs) are able to compete freely and fairly in the European single market.

In the era of cloud services, data sovereignty has become a cornerstone for assigning liability for the processing of the related data. The data sovereignty principle holds that the data protection rules and regulations of the country where the data is stored apply to the given processing entity. As such, larger companies such as Google, Amazon, or Facebook may be able to gain a competitive advantage by maintaining their servers in countries which are seen to serve their interests, while SMEs may not have the capacity to do the same.

This can have broad implications for the use of privacy-preserving technologies. If such a company perceives that the relevant data protection authority will be lenient in the case where the company did not implement privacy-preserving measures that resulted in the loss of personal data, they may be less inclined to implement those measures in order to avoid the associated costs. On the other hand, a smaller company that is subject to a stricter data protection authority may be more strongly encouraged to implement privacy-preserving measures in order to avoid fines that would more negatively impact their bottom line, and they may not have the legal teams to contest the fines nor be large enough to have a network effect to withstand the cost to their reputation. In fact, over the past several years, a number of firms declared bankruptcy in the UK due to the fines they received from the Information Commissioner's

[50] "Yes, Healthcare's Data Breach Problem Really Is That Bad", Healthcare Analytics News (25 September 2018), https://www.hcanews.com/news/yes-healthcares-data-breach-problem-really-is-that-bad.
[51] Ibid.
[52] Council Regulation (EC) No 1/2003 of 16 December 2002, OJ L1, 4.01.2003, Art. 23.
[53] For a more detailed analysis of how these provisions interact, see Deliverable 2.2.

Office.[54] While there are no doubt other concerns that factor into the equation in whether or not to use a given privacy-preserving technology (see section 3.1 below on costs and benefits), a company may not be largely affected due to a variety of factors, among them the network effects of holding a dominant position in the market.

One possible example of this may be seen in Facebook's Cambridge Analytica scandal. In this instance, Facebook improperly shared the data of up to 87 million users with Cambridge Analytica,[55] which was used to create psychological profiles of voters.[56] While there were two days of Congressional hearings on the matter in the US, with some members claiming that they were drafting new privacy-protecting laws, no new federal laws have been passed.[57] Apart from this, and despite the large scale of the data gathered, Facebook has received relatively little blowback. Within approximately two months after the details of the misuse of personal data surfaced, the company's stock had completely rebounded,[58] and the number of visitors to Facebook increased;[59] this may be due to the network effects of the platform and the lack of a competitive alternative.

As such, a more robust application of competition and antitrust law may be one method to ensure that these companies are implementing privacy-preserving technologies so that consumers are protected; flexibility in competition law is key in being able to address some of the challenges brought about by big data.[60]

2.4.3. Regional differences

There are a number of regional differences that must be kept in mind when discussing legal and ethical aspects that factor into the use of privacy-preserving technologies. For instance, the US does not have a comprehensive national privacy or data protection framework, and thus there exists no equivalent to the

[54] Tom Allen, "The ICO has only collected half of data breach fines since 2010", Computing (25 May 2018), https://www.computing.co.uk/ctg/news/3033019/the-ico-has-only-collected-half-of-data-breach-fines-since-2010.

[55] Cecilia Kang and Sheera Frenkel, "Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users", The New York Times (4 April 2018), https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html.

[56] Nicholas Confessore and Matthew Rosenberg, "Spy Contractor's Idea Helped Cambridge Analytica Harvest Facebook Data", The New York Times (27 March 2018), https://www.nytimes.com/2018/03/27/us/cambridge-analytica-palantir.html.

[57] Cecilia Kang and Kevin Roose, "Zuckerberg Faces Hostile Congress as Calls for Regulation Mount", The New York Times (11 April 2018), https://www.nytimes.com/2018/04/11/business/zuckerberg-facebook-congress.html.

[58] Jacob Sonenshine, "Facebook wipes out all of its losses following the Cambridge Analytica data scandal (FB)", Markets Insider (10 May 2018), https://markets.businessinsider.com/news/stocks/facebook-stock-price-wipes-out-cambridge-analytica-data-scandal-losses-2018-5-1024024416.

[59] Jake Kanter, "The backlash that never happened: New data shows people actually increased their Facebook usage after the Cambridge Analytica scandal", Business Insider (20 May 2018), https://www.businessinsider.com/people-increased-facebook-usage-after-cambridge-analytica-scandal-2018-5.

[60] N. Schepp and A. Wambach, *On Big Data and Its Relevance for Market Power Assessment*, Journal of European Competition Law & Practice, 2016, Vol. 7, No. 2, p. 123.

GDPR in the US.[61] There are, however, a variety of sectoral federal laws and regulations governing privacy and data protection in the US, for instance in relation to health or financial data.[62]

Schwartz & Janger stress the existence of differences between specific industries.[63] According to the authors, there are quite comprehensive regulations in the US, for instance, on financial institutions, whereas for retail and other non-financial entities, there are only guidelines of a non-binding nature.[64] For example, financial entities have to develop appropriate procedures for protecting customer data and to conduct periodic risk assessments. Another sector with relatively comprehensive regulations is healthcare.

In addition to federal law, there are also states that have enacted laws related to data protection. For instance, the State of California has one of the strongest legal privacy frameworks in the nation,[65] which includes a Constitution that "gives each citizen an 'inalienable right' to pursue and obtain 'privacy'" (Article 1) and has a series of privacy laws governing different sectors. California also enacted the first breach disclosure statute in the US, which went into effect in 2003.[66] Seemingly in the wake of the GDPR, there are a number of other states that have recently proposed bills that concern data protection, many of which create an obligation to report data breaches or to make the current notification requirement stricter.[67]

Some of these bills may also be in part a response to the aforementioned Equifax credit bureau breach that affected almost half the population of the United States, in addition to a large number of British and Canadian residents,[68] in order to make sure that data breaches are reported within a definite timeframe. Many states, such as Arizona, only had a requirement to report "in the most expedient manner possible

---

[61] The US privacy legislative framework has been outlined well by Canada's House of Commons Standing Committee on Access to Information, Privacy and Ethics. See Standing Committee on Access to Information, Privacy and Ethics, "Towards Privacy by Design: Review of the Personal Information Protection and Electronic Documents Act", pp. 70ff, http://www.ourcommons.ca/Content/Committee/421/ETHI/Reports/RP9690701/ethirp12/ethirp12-e.pdf

[62] Examples of federal legislation in the US that contains data protection rules in certain situations, includes: the Health Insurance Portability and Accountability Act (HIPAA) (42 U.S.C. §1301 et seq.), the Fair Credit Reporting Act (15 U.S.C. §1681), the Financial Services Modernization Act (Gramm-Leach-Bliley Act (GLB)) (15 U.S.C. §§6801-6827), the Electronic Communications Privacy Act (18 U.S.C. §2510), and the Controlling the Assault of Non-Solicited Pornography and Marketing Act (CAN-SPAM Act) (15 U.S.C. §§7701-7713 and 18 U.S.C. §1037) and the Telephone Consumer Protection Act (47 U.S.C. §227 et seq.).

[63] Paul M. Schwartz, Edward J. Janger, "Notification of Data Security Breaches", Michigan Law Review, Vol. 105, Iss. 5 (2007), https://repository.law.umich.edu/mlr/vol105/iss5/2.

[64] Ibid.

[65] A list of the major privacy protection laws at the state and federal level that are applicable in California may be found here: https://oag.ca.gov/privacy/privacy-laws.

[66] See California Security Breach Information Act, California Civil Code §§ 1798.29, 1798.82-.84. It should be noted that there is no time requirement in which to report the breach.

[67] Jeewon Kim Serrato, et al., "US states pass data protection laws on the heels of the GDPR", Data Protection Report (2018), https://www.dataprotectionreport.com/2018/07/u-s-states-pass-data-protection-laws-on-the-heels-of-the-gdpr/.

[68] Alex Hern, "Equifax told to inform Britons whether they are at risk after data breach", The Guardian (8 September 2017), https://www.theguardian.com/technology/2017/sep/08/equifax-told-to-inform-britons-whether-they-are-at-risk-after-data-breach.

and without unreasonable delay", and a bill has been introduced to require notification within 45 days.[69] However, despite a number of investigations into the breach, including one by the Federal Trade Commission (FTC), there has been no financial penalty imposed on Equifax.[70] While the company suffered a substantial loss in consumer trust,[71] their stock price has nearly returned to the level it had prior to the announcement of the breach.[72] Given the nature of the data that credit bureaus store, and the lack of consumer autonomy in relation to these credit bureaus, it is extremely important to have proper safeguards in place. As will be discussed below, the FTC is starting to acknowledge this and other similar issues in some of its recent hearings.

Chandler confirms that data breaches slightly affect the market value of companies.[73] However, breaches often seem to have more severe consequences for individuals than for the companies themselves. Although there is a growing number of cases, according to Chandler, chances are low that organisations are held accountable in the US or in Canada. It is not possible to prove negligence because it is difficult to establish that a breach caused identity fraud and, particularly if the identity fraud has not yet occurred, there is the problem of showing actual harm.

This situation may be contrasted with the situation in the EU,[74] where Article 32 of the GDPR imposes an obligation on data processing companies to implement adequate security measures in safeguarding the storage and processing of personal data, and Article 33 requires companies to report breaches within 72 hours; the violation of either provision can result in a penalty.[75] Some states in the EU also cover negligence: under Dutch law, there are also security requirements, and "serious culpable negligence" can result in an administrative fine and perhaps even a criminal procedure.[76]

There are also a number of associated costs for data breaches, including for detection and escalation, post data breach response, notification costs, and lost business cost.[77] For smaller scale data breaches,[78] the

---

[69] Privacy & Information Security Law Blog, "Arizona Amends Data Breach Notification Law" (24 May 2018), https://www.huntonprivacyblog.com/2018/05/24/arizona-amends-data-breach-notification-law/.

[70] Michael E. Kanell, "A year after data breach: Atlanta-based Equifax unbowed", The Atlanta Journal-Constitution (25 July 2018), https://www.ajc.com/business/year-after-data-breach-equifax-unbowed/YQVeBInUBd72EJwHWruQSK/.

[71] Ibid. A survey found that about half of respondents thought that Equifax should be banned from the credit bureau business.

[72] The Wall Street Journal, *Equifax, Inc. Stock Quote*, accessed 18 September 2018, https://quotes.wsj.com/EFX#.

[73] Jennifer A. Chandler, "Negligence Liability for Breaches of Data Security", Banking and Finance Law Review.

[74] For instance, the Information Commissioner's Office in the UK issued fines for data breaches even before the GDPR went into effect. Phil Muncaster, "ICO Fines Soared 69% in 2017", Infosecurity Magazine (26 January 2018), https://www.infosecurity-magazine.com/news/ico-fines-soared-69-in-2017/.

[75] GDPR, Art. 83.4(a).

[76] Dutch Data Protection Authority, "The data breach notification obligation as laid down in the Dutch Data Protection Act: Policy rules for the application of article 34a under the Dutch Data Protection Act", pp. 46-49, https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/policy_rules_data_breach_notification_obligation.pdf.

[77] Ponemon Institute, "2018 Cost of a Data Breach Study: Global Overview", p. 6, https://public.dhe.ibm.com/common/ssi/ecm/55/en/55017055usen/2018-global-codb-report_06271811_55017055USEN.pdf.

[78] Ibid. This study targeted data breaches of 100,000 records or less.

total associated costs have been shown to be higher in the US than in any other region or country.[79] Schwartz & Janger state that in the context of data breach notifications, regulatory forces meet with economic and reputation forces in the US.[80] As companies are under economic pressure to maximise profits, decision makers typically seek to calibrate expenditures according to the level of legal liability and the financial risks that they bear from leaked information. Moreover, they care about the reputational capital of their companies and seek both to avoid social sanctions and to gain social approval. This demonstrates how many of these issues are interrelated (for further details, see sections 3.2 and 3.3 on the aspects of "business models" and "public attention", respectively).

The aforementioned Federal Trade Commission is a federal agency with the mission to protect consumers and promote competition.[81] With respect to consumer protection, the FTC has the mission to stop unfair, deceptive and fraudulent business practices. In the US, the FTC is the main law enforcement actor at the federal level with regards to privacy. Primarily, the FTC enforces the Federal Trade Commission Act, which allows them to "prevent unfair methods of competition and unfair or deceptive acts or practices in or affecting commerce" and "seek monetary redress and other relief for conduct injurious to consumers.[82] The jurisdiction of the FTC with regards to privacy (and information security) arises from section 5 of the FTC Act, which prohibits unfair and deceptive methods, acts or practices in or affecting commerce.[83] The FTC also enforces specific-sector laws, such as the Children's Online Privacy Protection Rule (COPPA), the Safeguard Rule and the Fair Credit Reporting Act. Most of the FTC's privacy work is reactive. However, the FTC does provide some guidance for businesses on certain issues.[84]

In September 2018, the FTC held the second session of its 'Hearings on Competition and Consumer Protection in the 21st Century'.[85] Currently, the FTC uses the consumer welfare standard, and the hearing examined whether that standard is adequate today, and whether other public policy considerations, such as the size, wealth, or influence of corporations or individuals, income and wealth distribution, the bargaining power of large entities, or labour and employment considerations, should be taken into account in antitrust law.

[79] Ibid. See average total cost of data breaches, where the average cost of a data breach in the US was 7.91 million USD, compared with a worldwide average of 3.86 million USD. The latter figure represents an increase of 6.4 per cent since 2017.

[80] Paul M. Schwartz, Edward J. Janger, "Notification of Data Security Breaches", Michigan Law Review, Vol. 105, Iss. 5 (2007), https://repository.law.umich.edu/mlr/vol105/iss5/2.

[81] Federal Trade Commission, "About the FTC", https://www.ftc.gov/about-ftc.

[82] Federal Trade Commission, "Federal Trade Commission Act", https://www.ftc.gov/enforcement/statutes/federal-trade-commission-act.

[83] Federal Trade Commission, "Privacy and Security Enforcement", https://www.ftc.gov/news-events/media-resources/protecting-consumer-privacy/privacy-security-enforcement.

[84] See, e.g., Federal Trade Commission, "Advertising FAQ's: A Guide for Small Business", https://www.ftc.gov/tips-advice/business-center/guidance/advertising-faqs-guide-small-business.

[85] Federal Trade Commission, "FTC Hearing #2: Competition and Consumer Protection in the 21st Century", https://www.ftc.gov/news-events/events-calendar/2018/09/ftc-hearing-2-competition-consumer-protection-21st-century.

Recently, the FTC has been viewed as being lenient towards the enforcement of antitrust law, especially in comparison to the EU and other countries.[86] For instance, the EU fined Google 4.34 billion euros for ensuring that their search engine is dominant on Android mobile phones,[87] and India fined Google 18 billion euros for "unfair search bias".[88] As such, regional differences in the application of competition laws may result in varying levels of protection for consumers and of SMEs in doing business.

[86] Nitasha Tiku, "How to Curb Silicon Valley Power—Even With Weak Antitrust Laws", Wired (5 January 2018), https://www.wired.com/story/how-to-curb-silicon-valley-power-even-with-weak-antitrust-laws/.
[87] Jennifer Rankin, "Google fined £3.8bn by EU over Android antitrust violations", The Guardian (18 July 2018), https://www.theguardian.com/business/2018/jul/18/google-faces-record-multibillion-fine-from-eu-over-android.
[88] "Google fined by India watchdog for 'unfair search bias'", BBC News (9 February 2018), https://www.bbc.com/news/world-asia-india-42999519.

3. Societal and economic aspects

The assessment of existing technologies conducted within the scope of WP3 suggests that privacy-preserving technologies are integrated only to a limited extent in today's big data solutions. In our earlier deliverables, we demonstrated that there are quite strong solutions in research, but they are not widely implemented in existing big data solutions. Moreover, even if they are implemented, there is the risk for some of them that they are ignored or circumvented by the people using the big data solutions. In this section, we investigate the limited implementation. The aspects discussed are either considered to hamper the integration of privacy-preserving technologies into big data solutions or to play a role with respect to fostering their implementation. It is not our aim to suggest new technologies or to call for changes in existing technologies, it is to understand and eventually bring forward the implementation and use of the existing technologies in big data solutions. This does certainly not mean that there is no need or room for improvements on the technology side. The understanding of the implementation gap will be useful to develop design considerations for future big data solutions in the next phase of the project.

Currently, there is little knowledge on why existing privacy-preserving technologies are not more widely integrated into solutions and what could be done to change that. One explanation for the limited implementation, for instance, is suggested by Waldman[89] who conducted empirical research on how technology designers consider privacy in their operational processes and how law firm attorneys draft privacy notices. He argues that in most cases the main reason why privacy-preserving technologies do not seem to come from the ground is that outside technologists and lawyers have a very narrow understanding of privacy. In contrast to corporate privacy officers, who were found to push their companies to take consumer privacy seriously by integrating privacy into the designs of new technologies, outside technologists and lawyers often see privacy as limited to notice and synonymous with data security, respectively.

Jason Cronk, a US privacy lawyer and technologist, argues in an interview conducted by Solove[90] that from the perspective of companies practical reasons can also hamper the integration of privacy-preserving technologies. Some companies, according to Cronk, still reside in the early stages of privacy programme development. They, for instance, focus on developing high level policies and creating vendor management programmes. In Cronk's view, another problem underpinning the limited introduction of privacy-preserving technologies is that companies do not have working design processes. “They may have a development life-cycle for their core products or services, but even major companies sometimes lack formality in every aspect of their development.” In order to add formality into design processes, Cronk explains, companies tend to hire consulting.

Possible reasons for the limited integration of privacy-preserving technologies in big data solutions can be found on both sides, the supply side and the demand side of the market for big data solutions. The analysis conducted within the scope of the e-SIDES project focuses on developers and users of big data solutions,

[89] Ari Ezra Waldman, "Designing Without Privacy", (March 31, 2017). Houston Law Review, 55, https://ssrn.com/abstract=2944185

[90] Daniel Solove, "Strategic Privacy by Design: An Interview with Jason Cronk", https://teachprivacy.com/strategic-privacy-by-design/

the research community working on privacy-preserving technologies and data subjects. Insight into their views and actions may allow understanding why existing big data solutions are as they are and what affects decisions regarding the integration of privacy-preserving technologies into such solutions. We consider answers to these questions as essential for the development of reasonable design requirements for future big data solutions.

A central question when studying the integration of privacy-preserving technologies into big data solutions certainly is, if this makes sense economically (see section 3.1). One would expect that the handling of personal data as well as privacy protection are very important for developers and users of big data solutions, and data subjects alike. The assessment of existing privacy-preserving technologies, however, indicates that this is not always the case. Users seem to consider big data solutions that pay particular attention to privacy preservation as blocking, or in conflict with, their business models, whereas data subjects seem to be blinded by the benefits they get in return for their personal data (see section 3.2). At the same time, however, companies, including developers of big data solutions, increasingly seem to try to brand themselves as privacy protectors (see section 3.3). The fact that some data subjects are blinded by the benefits they get in return for their personal data becomes obvious when studying the economic value of privacy (see section 3.4). Moreover, when looking at the value of privacy, it becomes quite obvious that privacy-preserving technologies must be embedded in the products rather than provided as add-ons. More generally, is seems likely that cultural aspects determine the implementation and use of privacy-preserving technologies in the big data context to a significant extent (see section 3.5). Concerning people, in addition to culture, skills play a key role (see section 3.6). Policy makers and regulators definitely play an important role with respect to privacy preservation. Currently, apart from scandals reported by the media, legislation seems to be the key driver of related debates.

Within the EU, the GDPR requires all data processors to take protective measures (Art. 32 of the GDPR), including the application of privacy-preserving technologies. Consequently, it does not come as a surprise that societal and economic aspects related to the integration of technologies into big data solutions are closely related to legal and ethical aspects. As legal enforcement mechanisms lag behind in issuing fines or imposing other sanctions for companies suffering data breaches, incentives for companies to invest into comprehensive solutions diminish. As shown in section 2, even recurring incidents do not seem to trigger a rethinking. Similarly, privacy by design is regularly mentioned in related debates, and also in legal texts, as a promising and necessary approach to systems engineering, but it does not seem to be used comprehensively by developers of big data solutions. Ziegeldorf et al.[91] state that the economics of privacy are still in favour of those in disregard of privacy legislation. On the one side, development of privacy-enhancing technologies, enforcement, and audits of privacy protection policies is expensive and can limit business models. On the other side, violations of privacy legislations either go unpunished or result only in comparably small fines, whereas public awareness is still too low to induce intolerable damage of public reputation.

Based on desk research, the following aspects are discussed in this section. The aspects were selected based on previous work in the e-SIDES project, mainly the general assessment of privacy preserving

[91] Jan Henrik Ziegeldorf et al. "Privacy in the Internet of Things: threats and challenges", doi:10.1002/sec.795

technologies in WP3 (see section 4 of D3.2). Concrete examples from different application contexts are given to illustrate the points mentioned:

- **Costs and benefits**: Adding privacy-preserving technologies to big data solutions leads to additional costs for solution developers and users. These costs must be offset by the expected benefits. Key questions addressed include: Does the integration of privacy-preserving technologies in big data solutions make sense from an economic point of view? What are the actual costs and benefits to take into account?
- **Business models**: Putting privacy principles such as purpose limitation or data minimisation into practice may be in conflict with current or desired business models. Key questions addressed include: How does taking privacy preservation more seriously affect business models? To what extent does the protection of privacy lead to a reduction of the utility of data?
- **Public attention**: Taking privacy and transparency seriously, and making this public has the potential to allow for competitive differentiation. Key questions addressed include: Does privacy preservation have the potential to become a relevant differentiator for developers of big data solutions? How relevant is transparency in the context of big data?
- **Economic value**: Understanding the value of privacy for individuals is essential as this allows designing appropriate big data solutions. Individuals have the potential to exert significant pressure on actors in the data value chain. Key questions addressed include: How valuable is personal data from the perspective of data subjects? What affects the value of personal data in the era of big data?
- **Cultural fit**: Acknowledging that privacy preferences and practices vary among nations and regions, and are affected by cultural values is important. Key questions addressed include: How relevant are cultural aspects in the context of privacy? How do cultural values influence consumers' concerns about privacy?
- **Skill level**: Adapting to a new mind-set seems to be necessary as data has become a strategic business asset and privacy a threatened value. Key questions addressed include: Do developers of big data solutions have the skills required to transfer research results into products? What are the skills that are needed?

Obviously, there are considerable regional differences with respect to the use of data and big data solutions as well as the protection of privacy; particularly if looking at large economic blocks such as the US and the EU, they become obvious quickly. This is also suggested by the assessment of existing technologies conducted within the scope of WP3. In the US, for instance, it is common not to be too restrictive and to see what happens, and then generalise and apply case-based legal decisions. In contrast, the historical context in the EU is more rule-driven. To explain the limited integration of privacy-preserving technologies in today's big data solutions, it is critical not only to consider regional differences but also to understand what the differences mean at the global level. It is no secret that a significant share of widely-used big data solutions are developed by companies headquartered in the US.

Understanding why existing big data solutions are as they are requires clarity about relevant regional differences and their global impact. Moreover, reasonable design requirements may also have to be adjusted to regional differences. Regional differences are clearly relevant with respect to aspects such as

public perception, economic value of privacy and cultural fit. Bellman et al.[92], for instance, state that companies will increasingly have to customise their information collection and management strategies to match the privacy concerns of consumers in different regions. The authors found that most of these concerns are highly related to the privacy regulatory framework prevailing in a particular country, which tends to reflect as well as shape the privacy preferences of individual consumers. A sample of Internet users from 38 countries was matched against the Internet population of the US. Another influence on online privacy concerns is a lack of experience with the Internet. This influence, however, according to Bellman et al., is likely to diminish as the average level of experience grows. Finally, some of the observed differences in concern seem to reflect differences in cultural values that are likely to persist. They, however, only become apparent if privacy regulations across countries are harmonised.

Steinke[93] and Whitman[94], for instance, focus explicitly on differences between Europe or the EU on the one side, and the US on the other side. Steinke emphasises that there are different approaches to data privacy and protection. Whereas the US bets largely on self-regulation, there are rather strict legal requirements in Europe. The author stresses that even though privacy expectations and legal requirements differ depending on culture and government, customers generally prefer websites where a maximum of privacy protection is provided. If this privacy can be technically guaranteed, consumers, according to Steinke, would be even more supportive of those websites and organisations. Whitman concludes that there is little reason to suppose that Americans will be persuaded to reconsider their values and think in a European way any time soon. According to the author, US law simply does not endorse the general norm of personal dignity found in Europe, nor is there any hope that Europeans will embrace the American position. Whitman stresses that there is no such thing as privacy *as such*. The battle, if it is fought at all, has to be fought over more fundamental values than privacy.

3.1. Costs and benefits

Without doubt, the addition of privacy-preserving technologies to big data solutions leads to costs for the developers of the solutions. Acquisti et al.[95] clearly state that there are costs associated with the act of protecting data in general and the use of privacy-preserving technologies in particular. However, there is little knowledge about the amount of the costs involved. Less obvious is that the use of such solutions may also be more costly than the use of big data solutions without privacy-preserving technologies. Regarding benefits, there is no evidence that the offer of privacy-preserving big data solutions will lead to increased sales for developers or justify higher prices. Users of such solutions, provided that they actually use the privacy-preserving features, may benefit from lower costs related to data breaches.

[92] Steven Bellman, Eric J. Johnson, Stephen J. Kobrin, Gerald L. Lohse, "International Differences in Information Privacy Concerns: A Global Survey of Consumers", doi:10.1080/01972240490507956
[93] Gerhard Steinke, "Data privacy approaches from US and EU perspectives", doi:10.1016/S0736-5853(01)00013-2
[94] James Q. Whitman, "The Two Western Cultures of Privacy: Dignity versus Liberty", http://digitalcommons.law.yale.edu/fss_papers/649
[95] Alessandro Acquisti, Curtis Taylor, Liad Wagman, "The Economics of Privacy", doi:10.1257/jel.54.2.442

Findings from cost-benefit analyses focusing on security systems are to a certain extent applicable to the privacy context. Alese et al.[96], for instance, provide an overview of relevant costs and benefits. The authors differentiate not only between costs and benefits related to IT and non-IT impacts but also between proactive and reactive security strategies. Figure 2 summarises the findings of Alese et al.

Security Strategy	IT Impacts	Non-IT Impacts
Proactive	**Cost:** Cutting-edge hardware and software likely more expensive than well-established solutions.	**Cost:** User inconvenience.
	Cost: Information gathering, installation, debugging, and maintenance costs (labor).	**Benefit:** Regulatory and reputation benefits.
	Benefit: Decreased need for reactive labor.	**Benefit:** Fewer business Interruptions.
Reactive	**Cost:** Infrastructure (mostly labor) resources needed to respond quickly and effectively.	**Cost:** More events, and thus a likely increase in down time
	Cost: Resources (labor) needed to repair damaged systems and data.	**Cost:** Potential damage to reputation

Figure 2: Overview of costs and benefits (Alese et al.)

Regarding proactive strategies, Alese et al. stress that cutting-edge hardware and software is likely to be more expensive than well-established solutions. This also seems to hold for privacy-preserving big data solutions as compared to established solutions. Whereas the costs for information gathering, installation, debugging and maintenance may not be affected significantly by the integration of privacy-preserving features into big data solutions, the costs related to user inconvenience are affected. User inconvenience may be caused, for instance, by the pre-processing of data, restrictions with respect to the accessing of data and additional documentation needs. Decreased need for reactive labour as well as regulatory and reputation benefits are also relevant in the privacy context. Business interruptions are rather unlikely in relation with incidents involving privacy and other societal and ethical issues.

With respect to reactive strategies, the costs of resources needed to respond quickly and effectively are also relevant in the privacy context, however this is not the case for the costs related to the resources needed to repair damaged systems and data. Without doubt, quick and effective responses are needed after serious incidents related to privacy and other societal and ethical issues. Typically, however, neither

[96] Boniface K. Alese et al., "Cost-Benefit Analysis of Cyber-Security Systems", WCECS 2016

the systems nor the data is damaged within the scope of such incidents and thus it does not need to be repaired. It is very likely that a reactive strategy will lead to more incidents and thus higher costs for responding to them as well as higher costs for potential damage to reputation but, as already mentioned, business interruptions and downtimes are quite unlikely.

Alese et al. state that more than half of the organisations they studied employ both a proactive and a reactive strategy. A considerable number, however, indicated that they do not take proactive measures. The reasons they provided include

- Disruption of staff/lower productivity
- Expensive products
- High complexity and expenditure of time
- Difficulty of convincing the management
- Anticipated staff resistance

The factors are closely related and likely to also be relevant to some extent in the context of privacy. There are several reasons that could lead to an unfavourable cost-benefit ratio for using privacy-preserving big data solutions. Concerning developers of big data solutions, the factors suggest that users are not ready to pay for the additional efforts related to integrating privacy-preserving features into big data solutions. The difficulty of convincing the management of the need for such solutions and the anticipated staff resistance receives further attention in the section focusing on culture (see section 3.5). It remains to be checked, however, if the findings can actually be transferred to the context of privacy preservation.

Khokhar et al.[97] describe a cost-benefit analysis of privacy and data utility in the healthcare context. Thereby, they shed light on aspects related to user inconvenience through data pre-processing in a specific application context. The authors examine relevant cost factors associated with the value of anonymised data and the possible damage cost due to potential privacy breaches. The analytical cost model used by Khokhar et al. is shown in Figure 3

[97] Rashid H. Khokhar et al., "Quantifying the costs and benefits of privacy-preserving health data publishing", doi:10.1016/j.jbi.2014.04.012

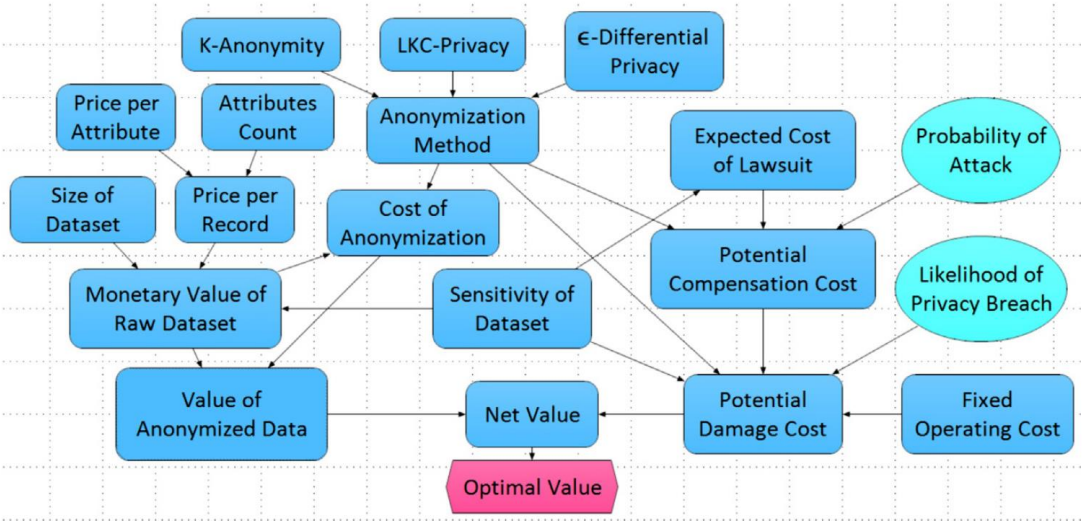

Figure 3: Analytical cost model (Khokhar et al.)

Factors that determine the value of anonymised data are the value of the raw dataset, which relies on the size and sensitivity of the dataset as well as the price per record, and the cost of anonymisation, which depends on the value of the raw dataset and the anonymisation method. The anonymisation method determines not only the cost of anonymisation in the sense of reduced utility, which may result from the removal of potentially relevant information, but also potential damages and compensation costs as it affects the risk of privacy breaches. The costs of potential damages include those incurred in sending mandatory breach notifications, dealing with regulatory investigations, hiring external auditors, facing class action litigation, and losing the goodwill of the general public due to decreased loyalty. Khokhar et al. aimed at providing guidance with respect to the selection of the optimal anonymisation method. Their findings, however, are also relevant to better understand the costs and benefits related to the inclusion of anonymisation features into big data solutions in particular, and privacy-preserving big data solutions in general.

With respect to healthcare, it is also interesting to see what data subjects think about the use of their data. By means of a survey, Trachtenbarg et al.[98] found that the best possible care is more important to patients than privacy. The authors surveyed more than 800 patients and 95% reported that they have no concerns about sharing medical information with their doctor and other medical personal. Trachtenbarg et al. also focused on costs. Slightly more than two-thirds of the survey participants indicated that they favour cost reduction over privacy. Only about one-third of participants were willing to spend more than five minutes on paperwork to improve privacy or to pay extra money for privacy (more details on the economic value of privacy are provided in section 3.4).

Nevertheless, it is still a challenge in healthcare that a single entity rarely holds all of the information necessary to conduct research or to provide clinical care. Patient health data is naturally spread across different sources due to visits at various medical practices, hospitals and pharmacies that often do not

[98] David E. Trachtenbarg, "The benefits, risks and costs of privacy: patient preferences and willingness to pay", doi:10.1080/03007995.2017.1292229

share the same information system or exchange data. According to Chen et al.[99], healthcare entities remain reluctant to exchange data due to privacy concerns. However, only linking patient-level data from various sources allows physicians and researchers to have more comprehensive data for clinical decision making and research. Iyengar et al.[100], for instance, also emphasise the need for both linking their medical records for better healthcare and the need for the protection of the patients' privacy.

According to Kulynych & Greely[101], researchers, providers and regulators must do more than aim to convince patients to accept the privacy risks of genomic research based on electronic medical records as an inescapable cost of receiving medical care. The research community's long practice of treating genomic information as de-identified or describing such data as 'anonymized' has impeded the development of community norms for data privacy and security in genomic research. Huang et al.[102] emphasise that for the large-scale application of personal genomics in research and clinical settings, it is not only essential to effectively address security and privacy issues but also to minimise storage costs. The costs must be justified by the benefits.

Another relevant application context is web surfing or, more specifically, the use of social networking sites. Vishwanath et al.[103], for instance, state based on a study with more than 500 college students that it appears that users' privacy management on Facebook is premised on the juxtaposition of the benefits of openness against the costs of maintaining privacy, where the focus is rather on the benefits than on the costs. Social need fulfilment (finding new friends, maintaining existing relationships and getting social support) was found to be the most significant benefit that influences how users manage their Facebook settings. Other benefits that drive Facebook use, such as the fulfilment of information and entertainment needs, seem to be less relevant, most likely because they do not require the user to divulge personal information. With respect to costs, the perceived severity and the perceived susceptibility of privacy incursions were found to have a significant impact on privacy management. Vishwanath et al. conclude that users focus solely on settings that they care about. According to the authors, knowing this allows for better security design and more effective communication about the benefits of various settings. Similarly, Min & Kim[104] found that the motivation of relationship management through social networking sites and their perceived usefulness for self-presentation lead users to disclose information. The results suggest that privacy concerns can be offset only by multiple beneficial factors. Apart from that, Min & Kim suppose that privacy-related policies and features are not of much help in making users of social networking sites

[99] Feng Chen et al., "Perfectly Secure and Efficient Two-Party Electronic-Health-Record Linkage", doi:10.1109/MIC.2018.112102542

[100] Arun Iyengar, "Healthcare Informatics and Privacy", doi:10.1109/MIC.2018.022021660

[101] Jennifer Kulynych, Henry T. Greely, " Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide ", doi:10.1093/jlb/lsw061

[102] Zhicong Huang et al. "A privacy-preserving solution for compressed storage and selective retrieval of genomic data", doi:10.1101/gr.206870.116

[103] Arun Vishwanath et al., "How People Protect Their Privacy on Facebook: A Cost-Benefit View", doi:10.1002/jasist.23894

[104] Jinyoung Min, Byoungsoo Kim, "How Are People Enticed to Disclose Personal Information Despite Privacy Concerns in Social Network Sites? The Calculus Between Benefit and Cost", doi:10.1002/asi.23206

feel more protected. This is a finding that should also be taken into account when trying to understand the integration of privacy features into big data solutions.

Car data is considered to have the potential to generate value through increased revenues, reduced mobility cost, and increased safety and security. Privacy, however, is deemed critical to enable new features and services from car data.[105] With respect to connected cars, Karnouskos & Kerschbaum[106] state that it is clearly cheaper not to implement privacy safeguards and let the various stakeholders use the data for whatever purpose they desire. The authors state clearly that not harvesting the benefits of cyber-physical systems such as connected cars is not a sustainable option. At the same time, however, the interest of consumers to maintain or control privacy needs to be respected. Hence, according to Karnouskos & Kerschbaum, a social debate needs to take place balancing the conflicting objectives between data use and privacy. Such discussions are not new, and are already ongoing, but they need to be considered for the context of specific applications contexts such as connected cars also. This debate will need to achieve a compromise by taking costs and benefits into account, and setting intended parameters for privacy-preserving technologies.

Articles focusing on the costs of privacy breaches are not only relevant with respect to assessing the cost-benefit ratio of integrating privacy features into big data solutions, but also with respect to the role of incidents in this context. Such articles allow one to better understand the costs related to potential damages to reputation (see section 2 for more details on the relevance of incidents).

3.2. Business models

Taking data protection seriously and putting foundational principles into practice may be in conflict with current or desired business models. Until the effective date of the GDPR, some companies might have accepted the risk of ignoring data protection compliance requirements. Given the extensive fines, which are now more likely to follow from non-compliance, more and more companies are forced to adopt relevant changes. Zarsky[107] discusses possible conflicts. The following deliberations are related to the ones in the section on privacy by design (see section 2).

Zarsky states that *purpose limitation* is clearly at odds with the prospect of big data analyses. Big data involves methods and usage patterns that neither the entity collecting the data nor the data subject considered or even imagined at the time of data collection. To comply with the purpose specification rule, entities striving to engage in big data analysis need to inform their data subjects of the future forms of processing they intend to engage in and closely monitor their practices to ensure they did not exceed the permitted realms of analyses. Carrying out any one of these tasks, according to Zarsky, might prove costly, difficult or even impossible.

[105] McKinsey & Company, “Car data: paving the way to value-creating mobility”, https://www.the-digital-insurer.com/wp-content/uploads/2016/05/704-mckinsey_car_data_march_2016.pdf
[106] Stamatis Karnouskos, Florian Kerschbaum, "Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles", doi:10.1109/JPROC.2017.2725339
[107] Tal. Z. Zarsky, "Incompatible: The GDPR in the Age of Big Data", https://scholarship.shu.edu/shlr/vol47/iss4/2

The GDPR defines *data minimisation*, stating that data must be "limited to what is necessary in relation to the purposes for which they are processed". The minimisation principle relates to the scope and categories of data initially collected as well as to the limited duration during which personal data may be retained. The rush towards big data, however, according to Zarsky, provides companies with a clear incentive to collect and retain as much data as they can for as long as possible. In theory, at least, with more data will come greater knowledge and thus greater benefit to the companies and potentially society in general.

Zarsky also describes conflicts related to different treatments of *special categories* of data. Special data categories include data "revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership and the processing of data concerning health or sex life". Whereas the justification for setting higher levels of protection for such data seems intuitive at first glance, enhanced forms of analytics challenge the ability to draw a distinction between special and other categories of data. An analysis merely relying on and addressing 'regular' categories of data can quickly end up pertaining to 'special' categories.[108]

The GDPR (Art. 22) sets forth a specific legal rule governing *decision-making processes*, which are both fully automated and substantially affect individuals, such as ones related to credit applications or recruiting. Individuals have the right not to be subjected to these processes. According to Zarsky, the rule signals a deep distrust towards automated processes but does not specify why this attitude was adopted. It is possible that companies will indeed be required to substantially change their technological architectures and even business models, opting for less efficient practices that comply with this rule.

Anderson & Stajano[109] focus on conflicts in the context of social networking. The authors state clearly that social networks have serious privacy drawbacks and that some of them are a direct consequence of the typical business model used. The authors argue that it is possible to develop alternative social networks that provide privacy and performance, while being both technically and economically viable. They suggest using an untrusted infrastructure together with local access control to ensure that users can control how the system shares their information. The authors conclude that privacy is not inherently incompatible with social networking.

Many business model conflicts result from the fact that there is a trade-off between privacy preservation and data use. This trade-off is particularly obvious with respect to anonymisation and sanitisation, but other approaches can also affect the use of data, for instance, by being inconvenient for the users of big data solutions. Iyengar et al.[110], for instance, stress that anonymisation and de-anonymisation of health data is well-addressed in regulatory compliance requirements, both in practice and in the scientific community, but challenges exist regarding the utility of data and the level of anonymization. According to Sharma et al.[111], the advantages of data-driven healthcare systems come with a price, an extraordinary

[108] Lokke Moerel, "GDPR conundrums: Processing special categories of data", https://iapp.org/news/a/gdpr-conundrums-processing-special-categories-of-data/
[109] Jonathan Anderson, Frank Stajano, "Must Social Networking Conflict with Privacy", doi:10.1109/MSP.2013.23
[110] Arun Iyengar, "Healthcare Informatics and Privacy", doi:10.1109/MIC.2018.022021660
[111] Sagar Sharma et al., "Toward Practical Privacy-Preserving Analytics for IoT and Cloud-Based Healthcare Systems", doi:10.1109/MIC.2018.112102519

effort to protect patients' privacy without compromising the utility of the data and associated healthcare services. Protecting privacy in healthcare systems is urgent and challenging because of the sensitivity of the associated data and the multifaceted exposure of healthcare systems. Homomorphic encryption and its variants might provide a solution in the future but they are far from being practical at this point.[112]

Chakravorty et al.[113] describe the same challenge for smart homes. The challenge, according to the authors, is to find the right trade-off between the amount of privacy and the loss of information. Mazhelis et al.[114] state that increased privacy may limit flexibility and innovation in the smart city context. They refer to research published by van de Kerk[115] in 2015. An analysis of four smart cities from the perspective of privacy protection revealed that only one actually fulfilled the EU regulations for personal data usage, but this limited flexibility and innovation compared to the other cities. With respect to connected cars, Karnouskos & Kerschbaum[116] state that one can have only either integrity or privacy. According to the authors, it is essential to effectively combine privacy-enhancing technologies with integrity-protecting mechanisms. Efforts bringing together privacy and integrity should not to be seen as an operational add-on, but as an integral part of the connected car life cycle.

According to Karnouskos & Kerschbaum[117], there is good scientific foundation to choose the key length in encryption. However, similar techniques are often missing for the choice of parameters in privacy-enhancing technologies. When using differential privacy, the choice of ϵ is difficult, and although partial progress has been made, even the choice of k in k-anonymity is still difficult.[118] Karnouskos & Kerschbaum consider more research into the implications and proper settings of those parameters useful and needed to help guide the concurrent debate.

Differences in their business models explain to some extent why it is more obvious for Apple to brand itself as a company that protects its customers' privacy than it is for other major technology companies such as Google, Facebook or Amazon. Apple's main business has been selling devices rather than advertising or e-commerce.[119] Apple will have to find the right balance as it intensifies its push into services as personalisation and thus insight into the individuals' preferences are essential in this context. Datta et al.[120], for instance, studied Google's ad privacy settings and found violations with respect to non-discrimination (i.e., protected attributes are not used for ad selection) and transparency (i.e., all data used for ad selection is visible). They found that the presence of protected attributes causes changes in ads as

[112] Arun Iyengar, "Healthcare Informatics and Privacy", doi:10.1109/MIC.2018.022021660
[113] Antorweep Chakravorty et al. "Privacy Preserving Data Analytics for Smart Homes", doi:10.1109/SPW.2013.22
[114] Oleksiy Mazhelis et al., "Towards enabling privacy preserving smart city apps", doi:10.1109/ISC2.2016.7580755
[115] Iris van de Kerk, "Data use versus privacy protection in public safety in smart cities", MSc thesis
[116] Stamatis Karnouskos, Florian Kerschbaum, "Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles", doi:10.1109/JPROC.2017.2725339
[117] Stamatis Karnouskos, Florian Kerschbaum, "Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles", doi:10.1109/JPROC.2017.2725339
[118] The parameters ϵ and k are used in differential privacy and k-anonymity, respectively, to quantify the privacy risk posed by releasing (statistics computed on) sensitive data.
[119] Julia Love, "Apple 'privacy czars' grapple with internal conflicts over user data", https://www.reuters.com/article/us-apple-encryption-privacy-insight-idUSKCN0WN0BO
[120] Amit Datta et al., "Automated Experiments on Ad Privacy Settings", doi:10.1515/popets-2015-0007

well as that attributes that are not in the settings cause changes in ads. The authors found no violations with respect to effective choice (i.e., changing a setting has an effect on ads) and ad choice (i.e., removing an interest decreases the number of ads related to that interest). As it is unlikely that the problems are simply the result of mistakes, a connection with Google's business model can be assumed.

3.3. Public attention

"Some major players are even changing their strategies in order to become leaders regarding data protection friendly products and services."[121] Privacy, security and data protection standards made in Europe seem to have the potential to become a trademark. Changes in the legislative framework play a key role with respect to this development. According to Acquisti et al.[122], being able to exploit privacy-friendly stances for competitive advantage is a clear benefit of not disclosing data for companies. Martin & Murphy[123] use the term 'privacy as a strategy' to refer to the phenomenon of using data protection approaches for competitive differentiation. The authors state that as long as companies compete in markets where measures for privacy protection can be differentiated and are valued by customers, using privacy as a strategy remains a viable option to marketers.

The Electronic Frontier Foundation (EFF) analyses on a yearly basis the policies and advocacy positions of major US technology companies when it comes to handing data to the government. The EFF states in its most recent 'Who has Your Back' report[124] that technology companies have increased transparency with respect to how and when they divulge data to the government. The EFF considers this shift to be fuelled in large part by public attention. According to the most recent report, every company that was evaluated had adopted baseline industry best practices such as publishing a transparency report and requiring a warrant before releasing user content to the government. Companies including Adobe, Dropbox, Pinterest, Uber and Wordpress received credit in all five categories.

Protecting consumer privacy met with considerable public interest with the refusal of Apple to grant US law enforcement backdoor access to the iPhone of a known terrorist. Indeed, headlines about the matter directly addressed the notion of using privacy as a strategy by referring to the government prosecutors' quotes that Apple's refusal "appears to be based on its concern for its business model and public brand marketing strategy"[125]. Beyond Apple, which seems to be willing to sacrifice some profit for the sake of privacy to bolster its image as a company that protects customers[126], companies are increasingly competing on the basis of strong consumer privacy protections. Using a privacy-centric approach to

[121] Jan Philipp Albrecht, "How the GDPR Will Change the World", doi:10.21552/EDPL/2016/3/4
[122] Alessandro Acquisti, Curtis Taylor, Liad Wagman, "The Economics of Privacy", doi:10.1257/jel.54.2.442
[123] Kelly D. Martin, Patrick E. Murphy, "The role of data privacy in marketing", doi:10.1007/s11747-016-0495-4
[124] Electronic Frontier Foundation, "Who Has Your Back? 2017: Protecting Your Data From Government Requests", https://www.eff.org/files/2017/07/08/whohasyourback_2017.pdf
[125] Eric Lichtblau, Matt Apuzzo, "Justice Department Calls Apple's Refusal to Unlock iPhone a 'Marketing Strategy'", https://www.nytimes.com/2016/02/20/business/justice-department-calls-apples-refusal-to-unlock-iphone-a-marketing-strategy.html
[126] Julia Love, "Apple 'privacy czars' grapple with internal conflicts over user data", https://www.reuters.com/article/us-apple-encryption-privacy-insight-idUSKCN0WN0BO

competitive differentiation, however, can be risky as experienced by Microsoft with Scroogled. The Scroogled campaign singled out Google as violating consumer privacy through practices such as scanning the contents of Gmail messages to personalise advertising and behavioural targeting. According to Martin & Murphy[127], a key reason for the campaign's failure was that Microsoft blamed Google for this approach instead of pointing out what Microsoft is doing right on the privacy front. Crampton[128] even states that documents show that several companies including Microsoft actually made it easier for the government to obtain data. These issues, according to Crampton, is very sensitive to the companies as they see themselves as competing on privacy. A key source of added value to their product is the personal privacy protection it supposedly offers. The documents appear to tell a very different story about privacy and geo-locational tracking.

Casadesus-Masanell & Hervas-Drane[129] state clearly that companies compete for consumer information and derive revenues both from consumer purchases as well as from disclosing consumer information in a secondary market. The authors conclude from their research that competition drives the provision of services with a low level of consumer information disclosure (high level of privacy). Moreover, however, they found that higher competition intensity in the marketplace does not necessarily improve privacy when consumers exhibit low willingness to pay. It is important to note, however, that Casadesus-Masanell & Hervas-Drane provide a benchmark for informed and rational consumers. Accordingly, the authors expect increasing consumer awareness of disclosure practices and familiarity with its implications to reinforce the relevance of their research. Section 3.4 shows that data subjects do not always act in a fully rational way in situations affecting their privacy.

Transparency, which goes well beyond notifications in case of data breaches, is a key aspect when discussing public attention in the big data context. Kulynych & Greely[130] state, focusing on research in the healthcare context, that the promise of big data and the appetite of researchers for access to information are enormous. In pursuit of new knowledge, participant consent in records-based research is increasingly abandoned, relying instead upon various degrees of de-identification to satisfy ethical concerns and meet regulatory requirements. According to Kulynych & Greely, there is very little transparency in most records-based research. Apart from blanket reassurances stating that 'your privacy is protected', providers don't offer patients specifics about who will receive what information.

Focusing on law enforcement, van Brakel[131] states that in addition to the increased responsibilities of technology companies, another problem is the fact that technology is often not transparent. From a business consideration, secrecy is important for companies developing such software, as being transparent about their algorithms may result in other companies potentially stealing their ideas. Van Brakel explains that a lot also depends here on who controls the algorithm. It is often unclear who provided the information on the subjects and how criteria were defined for algorithms to profile or risk

[127] Kelly D. Martin, Patrick E. Murphy, "The role of data privacy in marketing", doi:10.1007/s11747-016-0495-4
[128] Jeremy W. Crampton, "Collect it all: national security, Big Data and governance", doi: 10.1007/s10708-014-9598-y
[129] Ramon Casadesus-Masanell, Andres Hervas-Drane, "Competing with Privacy", doi:10.1287/mnsc.2014.2023
[130] Jennifer Kulynych, Henry T. Greely, " Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide ", doi:10.1093/jlb/lsw061
[131] Rosamunde van Brakel, "Pre-emptive big data surveillance and its (dis)empowering consequences: the case of predictive policing", in: Bart van der Sloot et al. (Eds.) "Exploring the Boundaries of Big Data"

assess these persons. This, according to her, is especially an issue with software for which hardly any alternatives are available. Moreover, there are usually no clear safeguards or redress procedures in place for people to be able to see their 'threat ratings' or to find out what indicators gave them a high score. There is no mechanism to correct errors, for instance, if someone has the same name as a convicted person.

3.4. Economic value

It is important to understand the economic value privacy and personal data has for individuals as this value affects the measures developers and users of big data solutions are willing to adopt to preserve privacy and address other ethical and societal issues. The higher the value, the more pressure data subjects will exert on the other actors in the data value chain[132]. The discussion of the value of privacy is related to the costs and benefits discussed in section 3.1.

The Organisation for Economic Co-operation and Development (OECD)[133] conducted a survey of methodologies for measuring the economic value of personal data. As shown in Figure 4, measures of the value of personal data are based either on market valuation or the valuation of individuals.

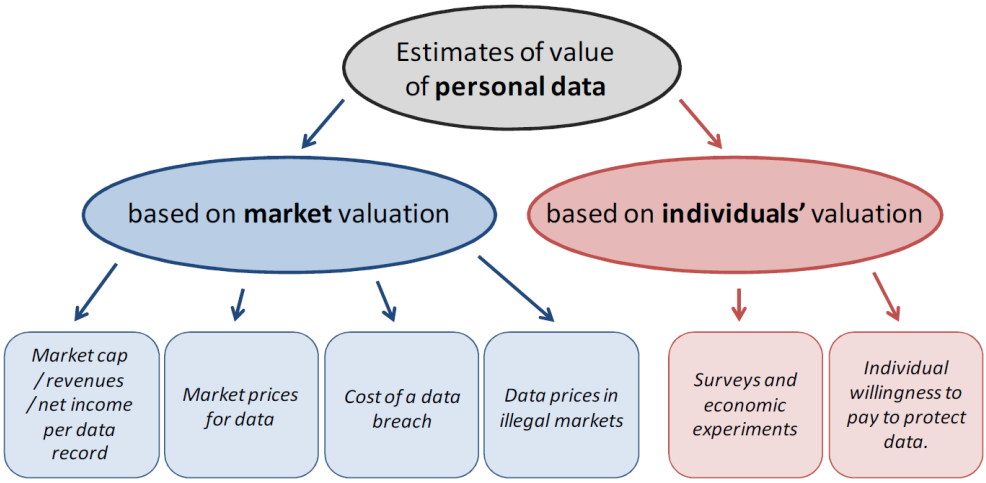

Figure 4: Estimates of value of personal data (OECD)

With respect to market valuation, measures can be based on financial results per data record, prices per personal data entry offered on the market by data brokers, the economic cost of a data breach per data entry and estimations of data prices in illegal markets. Measures relying on individual valuations can be based on valuations of personal data in monetary terms reported by individuals in surveys and economic experiments, as well as on the amounts that individuals are ready to spend to protect their personal data. Less conventional approaches, however, have also been pursued. In a Kickstarter campaign, for instance,

[132] Among the ones that come into question are data suppliers, technology providers, data end users, data marketplaces as well as regulators.

[133] OECD, "Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value", doi:10.1787/5k486qtxldmq-en

Federico Zannier decided to mine his own data to see how much it was worth[134]. He recorded all of his online activity, including the position of his mouse pointer and a webcam image of where he was looking, along with his GPS location data for $2 a day. In total, he raised over $2,700. Preibusch[135] reviewed several instruments and provides a guide to measuring privacy concerns. Among others, the author points out that hypothetical scenarios are not reliable and should not be used.

There is no shortage of economic literature attempting to quantify the value of data for both organisations and individuals[136]. According to Staiano et al.[137], such studies can be classified into two groups. The first and larger group includes studies that explicitly or implicitly measure the amount of money or benefit that a person considers to be enough to share their personal data, namely their willingness to accept (WTA) giving away their own data. The second and smaller group includes studies about tangible prices or intangible costs consumers are willing to pay (WTP) to protect their privacy. Hann et al.[138], for instance, quantify the value that individuals assign to protection against errors, improper access, and secondary uses of personal information online to an amount between $15 and $50. Olejnik et al.[139] found that elements of individuals' browsing histories are being traded among Internet advertising companies for amounts lower than $0.0005 per person.

The literature clearly shows that privacy concerns and expectations are remarkably context-dependent and very difficult to predict. Small changes in contexts and scenarios can lead to widely differing conclusions regarding the willingness to pay for access to or the protection of personal data. Carrascal et al.[140], for instance, conclude from their research that individuals value personal data related to their online identity significantly higher than data related to browsing activity. In a lab experiment, Tsai et al.[141] found that a substantial proportion of participants were willing to pay a small premium to purchase goods from merchants with more protective privacy policies. Jentzsch et al.[142], in contrast, found that only one third of participants were willing to pay a premium to purchase cinema tickets from a merchant that requests less personal information than a competing, but slightly cheaper, merchant. Moreover, the authors state, which is highly interesting in the e-SIDES context, that where there is a privacy difference between

[134] Joseph W. Jerome, "Buying and Selling Privacy: Big Data's Different Burdens and Benefits", Stanford Law Review Online

[135] Sören Preibusch, "Guide to measuring privacy concern: Review of survey and observational instruments", doi:10.1016/j.ijhcs.2013.09.002

[136] Alessandro Acquisti, Curtis Taylor, Liad Wagman, "The Economics of Privacy", doi:10.1257/jel.54.2.442

[137] Jacopo Staiano, Nuria Oliver, Bruno Lepri, Rodrigo de Oliveira, Michele Caraviello, Nicu Sebe, "MoneyWalks: A Human-Centric Study on the Economics of Personal Mobile Data", doi:10.1145/2632048.2632074

[138] Il-Horn Hann, Kai-Lung Hui, Tom Lee, I. Png, "Online Information Privacy: Measuring the Cost-Benefit Trade-Off", ICIS

[139] Lukasz Olejnik, Tran Minh-Dung, Claude Castelluccia, "Selling off privacy at auction", doi:10.14722/ndss.2014.23270

[140] Juan Pablo Carrascal, Christopher Riederer, Vijay Erramilli, Mauro Cherubini, Rodrigo de Oliveira, "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online", doi:10.1145/2488388.2488406

[141] Janice Y. Tsai, Serge Egelman, Lorrie Cranor, Alessandro Acquisti, "The effect of online privacy information on purchasing behavior", doi:10.1287/isre.1090.0260

[142] Nicola Jentzsch, Sören Preibusch, Andreas Harasser, "Study on monetising privacy: An economic model for pricing personal information", ENISA, https://www.enisa.europa.eu/publications/monetising-privacy/at_download/fullReport

companies, the privacy-unfriendly company typically obtains a greater market share. When the privacy-friendly company in one of the experiments of Jentzsch et al. charged €0.5 more (original ticket prices were about €7) than the privacy-invasive counterpart, its market share dropped from 90% to 42%. Similarly, Beresford et al.[143] found that a vast majority of participants chose to buy a DVD from a cheaper but more privacy-invasive merchant, than from a slightly costlier but less invasive merchant.

Acquisti et al.[144] explicitly challenge the premise that privacy valuations can be precisely estimated. They conclude form their research that privacy valuations are highly sensitive to non-normative influences. The authors show that privacy valuations are affected not only by endowment but also by the order in which different privacy options are described. Acquisti et al. found that order effects are weaker than endowment effects. With respect to psychological biases and heuristics that influence decisions regarding privacy even if all the necessary information was accessible, Brandimarte & Acquisti[145] list ambiguity aversion, self-control problems, hyperbolic time discounting, illusion of control, optimism bias and default bias. In addition to endowment effects and hyperbolic time discounting, van Lieshout[146] stresses that people perceive losses differently than gains and are more willing to prevent a loss than achieve a similar gain, are risk averse, tend to overvalue immediate rewards and undervalue long term rewards, tend to mimic behaviour shown by predecessors, and behave differently in the absence of real choices (e.g., getting access to a service requires accepting specific conditions).

Moreover, the value of privacy seems to depend on the social class to which an individual belongs. Higher class individuals are likely to be more educated (see section 3.6 for more information of skills) and thus empowered in dealing with consent mechanisms and the release of their data, and have more financial resources for court cases in case of privacy breaches. Moreover, they could afford to pay monthly fees for special privacy protection tools or features. Jerome[147] states that big data is all about categorisation and that categorisation threatens to place a privacy squeeze on the middle class as well as on the poor. Big data could, according to the author, lead to a democratisation of information but generally information is a more potent tool in the hands of the powerful. Categorisation may even be a more severe problem than data breaches, further favouring the rich. Similarly, Elvy[148] states that disadvantaged individuals are more prone to using privacy-invasive settings, packages and tools, thus revealing more information, which may be used for specific targeting, preying and discriminatory actions. Elvy even sees the emergence of an Internet class society that distinguishes between those who can afford to pay for more privacy and those who cannot.

---

[143] Alastair R. Beresford, Dorothea Kübler, Sören Preibusch, "Unwillingness to pay for privacy: A field experiment", doi:10.1016/j.econlet.2012.04.077

[144] Alessandro Acquisti, Leslie K. John, George Loewenstein, "What is Privacy Worth?", doi:10.1086/671754

[145] Laura Brandimarte, Alessandro Acquisti, "The Economics of Privacy", doi: 10.1093/oxfordhb/9780195397840.013.0020

[146] Marc van Lieshout, "The Value of Personal Data" doi:10.1007/978-3-319-18621-4_3

[147] Joseph W. Jerome, "Buying and Selling Privacy: Big Data's Different Burdens and Benefits", Stanford Law Review Online

[148] Stacy-Ann Elvy, "Paying for Privacy and the Personal Data Economy", Columbia Law Review

According to Crabtree et al.[149], the biggest current problem in the data economy is that individuals cannot reap benefits of the value of their data. The authors state that data subjects must be turned from victims to active economic players. Other than that, it is highly problematic that the trade of personal data is for most individuals a secondary, mostly inconspicuous and often altogether invisible aspect of a different, more salient transaction.[150] Crabtree et al. highlight that there are many factors that have an influence on the dichotomy between privacy attitudes and actual behaviour. This dichotomy is also known as the 'privacy paradox'. Finally, Schwartz & Solove[151] ask how different parties should share the wealth that the trade with personal data creates. Ideally, according to them, a market economy would permit the free price mechanism to set a price for the data. The authors argue that the lack of transparency regarding practices of data collection and tracking creates an asymmetry of knowledge about existing information collection practices between consumers and the organisations that collect information about them. This information asymmetry places consumers at a profound disadvantage in negotiations, such as they may exist, with those who collect their information. In sum, consumer objections to behavioural advertising are real and deserve a policy response.

3.5. Cultural fit

According to Hofstede[152], a particular nation exhibits cultural characteristics associated with particular sets of shared beliefs. Similar phenomena can also be observed for organisations. Culture provides people with sets of acceptable behaviours and norms within a group or society.[153] Global privacy research, including studies of how privacy preferences and practices vary among nations and regions, remains underdeveloped.[154] The limited findings show that cultural values can influence people's privacy perceptions such that countries with tighter privacy regulations experience fewer privacy problems.[155] Chakravorty et al.[156] emphasise that the concept of privacy varies from countries, cultures and jurisdiction. However in general, privacy is associated with collection, storage, use, processing, sharing or destruction of personally identifiable data.

Park[157] studied the congruence between policy supply and demand in Internet privacy as moderated by culture. He found that culture plays a significant role in creating the incongruence of the consensus

[149] Andy Crabtree, Tom Lodge, James Colley, Chris Greenhalgh, Richard Mortier, Hamed Haddadi, "Enabling the new economic actor: data protection, the digital economy, and the Databox", doi:10.1007/s00779-016-0939-3
[150] Alessandro Acquisti, Curtis Taylor, Liad Wagman, "The Economics of Privacy", doi:10.1257/jel.54.2.442
[151] Paul M. Schwartz, Daniel J. Solove, " The PII Problem: Privacy and a New Concept of Personally Identifiable Information", New York University Law Review 86
[152] Geert Hofstede, "Culture's Consequences: International Differences in Work-related Values", SAGE Publications
[153] Kyong-Jee Kim, Curtis J. Bonk, "Cross-cultural comparisons of online collaboration", doi:10.1111/j.1083-6101.2002.tb00163.x
[154] Kelly D. Martin, Patrick E. Murphy, "The role of data privacy in marketing", doi:10.1007/s11747-016-0495-4
[155] Sara Dolnicar, Yolanda Jordaan, "A Market-Oriented Approach to Responsibly Managing Information Privacy Concerns in Direct Marketing", doi:10.2753/JOA0091-3367360209
[156] Antorweep Chakravorty et al. "Privacy Preserving Data Analytics for Smart Homes", doi:10.1109/SPW.2013.22
[157] Yong Jin Park, "Privacy regime, culture and user practices in the cyber-marketplace", doi:10.1108/14636690810862811

between policymakers and actual users of different nations. Park concludes that it very much depends on cultural aspects whether or not marketplaces alone are able to take care of human rights concerns. It is not uncommon that top-down policy imposition occurs without due negotiation between policymakers and actual users.

Millberg et al.[158] found that a country's regulatory approach to the corporate management of information privacy is affected by its cultural values and by individuals' information privacy concerns. They also found that most companies take a primarily reactive approach to managing privacy by waiting for an external threat before crafting cohesive policies that confront their information practices. However, consumers and legislators in different societies exhibit varying levels of concern about information privacy, both in general and in their assessment of specific practices. Therefore, according to Millberg et al., a universal regulatory approach to information privacy seems unlikely and would ignore cultural and societal differences. The authors conclude that the self-regulatory model of privacy governance would not have been sustainable over the long term.

Bellman et al.[159] found that cultural values have an influence on consumers' concerns about information privacy. Cultural values were found to affect two dimensions of information privacy concerns: errors in databases and unauthorised secondary use. More specifically, power distance, individualism and masculinity were found to have an impact. The influence of uncertainty avoidance was not significant in the study. According to Debatin et al.[160], safer use of social network sites would require a dramatic change in user attitudes. A responsible and informed user would be required with a high level of computer literacy, not only in the technical but also in the socio-cultural and ethical sense. Dunnavant & Childress[161] explain, for instance, how a university evolved its culture to develop better habits respecting data security and privacy. Millberg et al.[162] state clearly that 'one size does not fit all' with respect to regulatory implementation.

Mooney & Pejaver[163] consider the ethical implications of the big data revolution with particular emphasis on maintaining appropriate care for privacy in a world in which technology is rapidly changing social norms regarding the need for privacy. With respect to public health research and practice, the authors see three key issues that big data raises: the risk of inadvertent disclosure of personally identifying information, the potential for increasing dimensionality of data to make it difficult to determine if a data set is sufficiently de-identified to prevent deductive disclosure of personally identifying information, and the challenge of

[158] Sandra J. Milberg, H. Jeff Smith, Sandra J. Burke, "Information Privacy: Corporate Management and National Regulation", doi:10.1287/orsc.11.1.35.12567

[159] Steven Bellman, Eric J. Johnson, Stephen J. Kobrin, Gerald L. Lohse, "International Differences in Information Privacy Concerns: A Global Survey of Consumers", doi:10.1080/01972240490507956

[160] Bernhard Debatin, Jennette P. Lovejoy, Ann-Kathrin Horn, Brittany N. Hughes, "Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences", doi:10.1111/j.1083-6101.2009.01494.x

[161] Susan Dunnavant, M. Jean Childress, "A Sideways Approach to Data Security and Privacy Awareness", doi:10.1145/1878335.1878384

[162] Sandra J. Milberg, H. Jeff Smith, Sandra J. Burke, "Information Privacy: Corporate Management and National Regulation", doi:10.1287/orsc.11.1.35.12567

[163] Stephen J. Mooney, Vikas Pejaver, "Big Data in Public Health: Terminology, Machine Learning, and Privacy ", doi:10.1146/annurev-publhealth-040617-014208

identifying and maintaining standards of ethical research in the face of emerging technologies that may shift the generally accepted norms regarding privacy.

A key question is whether big data rather leads to de-individualisation and discrimination, or personalisation. Van Wel and Royakkers[164], for instance, state that anonymous profiling could be harmful, and lead to possible discrimination and de-individualisation. People could be discriminated against, or become stigmatised simply by being labelled as a member of a group or by being labelled as an individual with certain characteristics. Some criteria, like race and religion, can be inappropriate and discriminatory to use in decision-making. Individuals can no longer protect themselves with traditional privacy rules and people could be judged, treated, and possibly discriminated against or stigmatised based on characteristics they do not even possess. According to van Wel and Royakkers, web data mining jeopardises different values like privacy and individuality, and other underlying values like non-discrimination, fair judgement and fair treatment. One of the main goals of web data mining, as used in e-commerce, is indeed personalisation: customising web sites and services to the individual wishes of each visitor. So, it does appear to promote individuality instead of threatening it. If, however, the personalisation is done by creating non-distributive group profiles, it can lead to de-individualisation and discrimination.

Nixon et al.[165] state that anonymity is a trade-off in the privacy debate as it may inhibit one of the core aims of a smart environment, which is to provide personalised interactions. According to Hernández-Ramos et al.[166], people demand richer experiences through more customised services being provided by any smart object. Hernández-Ramos state that smart environments pose conflicting requirements that must be adequately regulated.

3.6. Skill level

Data has become a strategic business asset. Failures in data security and governance regularly create public embarrassments for companies.[167] Therefore, it is essential that the big data and analytics skills discussion does not only focus on the role of the data scientist but on the fact that every professional occupation must adapt to the new mind-set. With respect to big data professionals, Miller[168] states that they do not only need to have knowledge in math and statistics, machine learning, predictive analytics, decision management, computer science and programming, but also data ethics, information law, information privacy, data security, data and information theory, and visualisation and communication skills in addition to the core information systems, database, data warehousing and data mining skills most

[164] Lita van Wel, Lambèr Royakkers, "Ethical issues in web data mining", doi:10.1023/B:ETIN.0000047476.05912.3d
[165] Patrick A. Nixon et al., "Security, Privacy and Trust Issues in Smart Environments", doi:10.1002/047168659X.ch11
[166] José L. Hernández-Ramos et al., "Preserving Smart Objects Privacy through Anonymous and Accountable Access Control for a M2M-Enabled Internet of Things", doi:10.3390/s150715611
[167] Charles Duhigg, "How Companies Learn Your Secrets", https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html
[168] Steven Miller, "Collaborative Approaches Needed to Close the Big Data Skills Gap", doi:10.7146/jod.9823

professionals already have today. Several of these skills are essential for developers to make sure privacy-preserving features are properly integrated into their big data solutions.

Mooney & Pejaver[169] also stress that the use of big data in public health research and practice calls for new skills to manage and analyse these data, though it does not remove the need for the skills traditionally considered part of public health training, such as statistical principles, communication, domain knowledge, and leadership. According to the Ponemon Institute[170], healthcare organizations continue to depend mainly upon policies and expertise to respond to data breaches. At least in healthcare, technologies appear to be considered as less effective in this regard. Whereas triability[171] and compatibility are certainly achievable with respect to privacy-preserving technologies, the lack of observability, a high complexity, and the missing or unclear relative advantage, may remain significant barriers to the adoption of such technologies.

Mazhelis et al.[172] argue that if privacy can be guaranteed to an acceptable degree on the level of the technical platform in a way, which prevents privacy breaches on the application level, it may be easier for application developers to focus on innovation as they have less responsibility on the privacy design. According to the authors, privacy controls are normally provided in platforms by means of confinement mechanisms that allow the users to explicitly define how much privacy they are ready to waive in exchange for desired functionality. Unfortunately, such mechanisms are often misused by application providers, and the users themselves are not always capable of assessing the privacy implications of their decisions.

Skills, however, are not only an important topic for developers but also for users of big data solutions and data subjects. In a study conducted to examine mobile-based privacy literacy by Park & Jang[173], less than half of the interviewed individuals possessed basic information and location privacy knowledge, privacy skills, and awareness of the risks associated with commercial mobile environments. Moreover, the authors found that a high level of mobile familiarity does not translate into knowledge as frequent mobile use was not associated with privacy knowledge and skill. Privacy aspects linked to mobility and social networks are getting increasingly relevant in the context of web surfing.

Bartsch & Dienlin[174] studied privacy literacy in the context of social networking sites. They found that time spent on Facebook and experience with privacy regulation do not per se increase safety and privacy behaviour directly. The authors stress that this underlines the importance of online privacy literacy as a mediator to a safe and privacy-enhancing online behaviour. Bruce & Lee[175] state that whereas teenagers

[169] Stephen J. Mooney, Vikas Pejaver, "Big Data in Public Health: Terminology, Machine Learning, and Privacy ", doi:10.1146/annurev-publhealth-040617-014208
[170] Ponemon Institute, "Sixth Annual Benchmark Study on Privacy & Security of Healthcare Data"
[171] Rossnagel defines 'triability' as the degree to which experimentation an innovation can be experimented with.
[172] Oleksiy Mazhelis et al., "Towards enabling privacy preserving smart city apps", doi:10.1109/ISC2.2016.7580755
[173] Yong Jin Park, S. Mo Jang, "Understanding privacy knowledge and skill in mobile communication", doi:10.1016/j.chb.2014.05.041
[174] Miriam Bartsch, Tobias Dienlin, "Control your Facebook: An analysis of online privacy literacy", doi:10.1016/j.chb.2015.11.022
[175] Ndibanje Bruce, Hoon Jae Lee, "Anonymization Algorithm for Security and Confidentiality of Health Data Set across Social Network", doi:10.1109/ICTC.2014.6983085

might have a clear understanding of privacy and related issues, this does not apply to some adults who have not even used e-mail and other basic Internet services before the social networking revolution, who not only have limited computer-related technical skills but also lack risk consciousness about privacy issues. Focusing on anonymity services, Rossnagel[176] points out that potential adopters are neither aware of the privacy risks they face when communicating online nor are they aware of the available technology that can protect themselves against these risks.

According to Kshetri[177], big data is likely to affect the welfare of unsophisticated, vulnerable and technologically unsavvy consumers more negatively than others. Such consumers may lack awareness of multiple information sources and are less likely to receive up-to-date and accurate information about multiple suppliers in a manner that facilitates effective search and comparisons. They are also not in a position to assess the degree of sensitiveness of their online actions and are more likely to be tricked by illicit actors. Further social class effects are discussed in section 3.4. Digital literacy is recognised as a key skill in the 21st century.[178] For all consumers to be protected, they must develop an understanding of how data can be collected and shared, how companies analyse data and for which purposes it is used. According to Debatin et al.[179], people need to be educated about risks to their privacy in a way that actually alters their behaviour.

[176] Heiko Rossnagel, "The Market Failure of Anonymity Services", doi:10.1007/978-3-642-12368-9_28
[177] Nir Kshetri, "Big data's impact on privacy, security and consumer welfare", doi:10.1016/j.telpol.2014.10.002
[178] L. H. Segura Anaya, Abeer Alsadoon, N. Costadopoulos, P. W. C. Prasad, "Ethical Implications of User Perceptions of Wearable Devices", doi:10.1007/s11948-017-9872-8
[179] Bernhard Debatin, Jennette P. Lovejoy, Ann-Kathrin Horn, Brittany N. Hughes, "Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences", doi:10.1111/j.1083-6101.2009.01494.x

4. Limits of technological approaches

There is wide consensus that a combination of technical and non-technical measures is essential for the privacy-preserving use of data. The Cambridge Analytica and Facebook scandal,[180] for instance, shows vividly the limitations of reactive approaches that prescribe norms and actions that are taken if there is a breach of privacy or some rules are broken. Technical solutions that are proactive in the sense that they prevent breaches or rule violations in the first place are required. Technological solutions alone, however, cannot yet make sure that contexts are properly taken into account. The assessment of existing privacy-preserving technologies supports the view that technical and non-technical measures need to be combined. However, it remains unclear how this combination is best realised, particularly as the technology landscape continuously evolves.

Technical and non-technical solutions are considered as equally important, complementary and partly overlapping with respect to their potential. Consequently, insight into the role that non-technical measures can play is essential not only for developing reasonable design requirements for future big data solutions, but also for better understanding the design of existing big data solutions. Many of the aspects discussed in sections 2 and 3 are highly relevant with respect to the combination of technical and non-technical measures.

Borking & Raab,[181] for instance, state that privacy-enhancing technologies are being encouraged in many countries, as part of a comprehensive and systematic approach to privacy protection that accords a significant role to technological means of protection without assuming that they are a 'magic bullet' that can be aimed at the target without the accompaniment of legal, organisational, ethical and educational tools. Moreover, the authors stress that it needs to be checked via privacy auditing or specific privacy-enhancing technology scans, whether equipped systems indeed comply with privacy legislation. Finally, Borking & Raab refer to certifications that might help in offering the necessary certainty to citizens and consumers that their privacy is being effectively protected. According to Clarke[182], effective protection is dependent on a multi-partite, tiered framework, in which layers of technology, organisational practices and law combine to ensure reasonable behaviour. Van Wel and Royakkers[183] stress that privacy policies are often difficult to understand, hard to find, take a long time to read, and can be changed without notice. Clearly, according to them, regulation and self-regulation efforts do not offer sufficient protection for web users' privacy.

Across application contexts, privacy-preserving technologies as well as big data solutions have weaknesses that still need to be addressed:

---

[180] Cecilia Kang and Sheera Frenkel, "Facebook Says Cambridge Analytica Harvested Data of Up to 87 Million Users", The New York Times (4 April 2018), https://www.nytimes.com/2018/04/04/technology/mark-zuckerberg-testify-congress.html

[181] John J. Borking, Charles D. Raab, "Laws, PETs and Other Technologies for Privacy Protection", http://www.egov.ufsc.br/portal/sites/default/files/anexos/27156-27166-1-PB.pdf

[182] Roger Clarke, "Platform for Privacy Preferences: A Critique", Privacy Law & Policy Reporter, 5(3)

[183] Lita van Wel, Lambèr Royakkers, "Ethical issues in web data mining", doi:10.1023/B:ETIN.0000047476.05912.3d

Mooney & Pejaver[184] stress that appropriate use of both big data and machine learning relies on understanding several key limitations of each. First, machine learning's capacity to overcome the curse of dimensionality requires large data sets. Small or biased training sets can lead to overfitting, which limits the problems that current machine-learning methods can address. Second, machine-learning models are often described as 'black boxes' whose opacity precludes interpretability or sanity-checking of key assumptions by non-experts. Although recent work, according to Mooney & Pejaver, has partially addressed this limitation, the problem persists. Third, in some instances, observers assume that models that learn automatically from data are more objective and therefore more accurate than human-constructed models. Although data-driven models can frequently predict outcomes better than theory-driven models, data-driven model building also involves subjective decisions, such as choice of training and evaluation data sets, choice of pre-processing criteria, and choice of learning algorithms and initial parameters. These decisions cumulatively result in biases and prejudices that may be obscured from casual users. Fourth, data quantity often comes at the expense of quality. This is an issue for any big data analysis but may be especially pernicious in the context of machine-learning methods that use a test set to estimate prediction accuracy in the broader world. If data collection artefacts render training and test sets overly similar to each other but different from those of the data sets to which the model would typically be applied, overfitting may lead to unanticipatedly poor prediction accuracy in the real world. Finally, because big data studies often require linking secondary-use data from heterogeneous sources, discrepancies between these data sources can induce biases, including demographically patterned bias.

Because genomic information is highly personal, privacy has become a major concern as this data become more widely generated, disseminated, and unwillingly exposed. For example, coarse-grained encryption and access control to genomic data, according to Huang et al.[185], could lead to incidental findings that doctors often prefer to avoid. Standard sample de-identification has been proven insufficient for complete protection of genetic privacy. Establishing a secure and privacy-preserving solution for genomic data storage is urgently needed to facilitate the usage and transfer of sequence data. For example, storing sequence data on a cloud is an attractive option, considering the size and the required availability of the data. However, access threats in this case are even more serious because the data owner has to trust insiders on the cloud (e.g., the cloud administrator or high-privileged system software). There is thus a need to integrate encryption methods into compression solutions for genomic data that are secure and privacy-preserving. The closest example of such a solution provides a privacy-preserving solution for storing BAM files,[186] but it does not provide an efficient method for compression.

Similar problems were described for the banking and finance context. According to de Montjoye et al.,[187] coarsening data is not enough to protect the privacy of individuals in financial metadata data sets. Although unicity decreases with the resolution of the data, it only decreases slowly regarding the

[184] Stephen J. Mooney, Vikas Pejaver, "Big Data in Public Health: Terminology, Machine Learning, and Privacy ", doi:10.1146/annurev-publhealth-040617-014208

[185] Zhicong Huang et al. "A privacy-preserving solution for compressed storage and selective retrieval of genomic data", doi:10.1101/gr.206870.116

[186] BAM is a binary format for storing sequence data.

[187] Yves-Alexandre de Montjoye et al. "Unique in the shopping mall: On the reidentifiability of credit card metadata", doi:10.1126/science.1256297

perspectives space, time and price. Furthermore, this decrease is easily overcome by collecting a few more points. For instance, at a very low resolution of 15 days, 350 shops and an approximate price 0.5, there is less than a 15% chance of re-identifying an individual knowing four points. However, if 10 points are known, there is more than an 80% chance of re-identifying a person. Montjoye et al. stress that this means that even noisy or coarse financial data sets along all of the dimensions provide little anonymity.

Inadequate user control seems to be a key issue in the web surfing context. Current social networks and photo-sharing sites mainly focus on the privacy of users' own media in terms of access control, but do little to deal with the privacy implications created by other users' media. According to Smith et al.[188], the issue of staying on top of what others are uploading, that might also be relevant to the user, is still very much outside the control of that user. So far, where privacy is concerned there has been a lot of work in the small data area (i.e., how can users control who has access to what they post themselves), but the big data issue in this area has focused almost entirely on what the controlling companies do with this information.

Hernández-Ramos et al.[189] stress that there is a lack of mature and suitable approaches for privacy preservation in smart environment. The main objective of privacy preservation is ensuring that private data remains protected, while processing or releasing sensitive information. Privacy concerns about data from smart homes have been raised in various publications. However, according to Chakravorty et al.,[190] there has been little work, on the design of technical solutions protecting privacy throughout the complete data lifecycle for smart home analytics. All of the discussed solutions are very narrow and address privacy concerns required for specific solutions. None of them can be easily incorporated or extended to existing or new smart home designs having different data processing needs. A key challenge is, according to Karnouskos & Kerschbaum,[191] to find a way to allow harvesting the benefits of added value services on the one side and prevent unwanted inferences on the other side. Sicari et al.[192] points out that traditional security countermeasures and privacy enforcement cannot be directly applied to all devices in smart environments due to their limited computing power; moreover the high number of interconnected devices arises scalability issues. At the same time, to reach full acceptance by users, according to Sicari et al., it is mandatory to define valid security, privacy and trust models suitable for smart environments.

Finally, in the law enforcement context, a key issue is lack of transparency. Big data solutions, such as pre-emptive surveillance technologies, according to van Brakel,[193] make it more difficult to scrutinize discriminatory effects as the technology and the aggregate level of analysis create an appearance of objectivity. What is specific for big data, however, is that the more different types of data are included in

[188] Matthew Smith et al. "Big Data Privacy Issues in Public Social Media", doi:10.1109/DEST.2012.6227909
[189] José L. Hernández-Ramos et al., "Preserving Smart Objects Privacy through Anonymous and Accountable Access Control for a M2M-Enabled Internet of Things", doi:10.3390/s150715611
[190] Antorweep Chakravorty et al. "Privacy Preserving Data Analytics for Smart Homes", doi:10.1109/SPW.2013.22
[191] Stamatis Karnouskos, Florian Kerschbaum, "Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles", doi:10.1109/JPROC.2017.2725339
[192] Sabrina Sicari et al., "Security, privacy and trust in Internet of Things: The road ahead", doi:10.1016/j.comnet.2014.11.008
[193] Rosamunde van Brakel, "Pre-emptive big data surveillance and its (dis)empowering consequences: the case of predictive policing", in: Bart van der Sloot et al. (Eds.) "Exploring the Boundaries of Big Data"

the analysis, the greater the likelihood of bias. De Hert & Lammerant[194] stress that people, including state officials, are often fooled by technology or do not possess the ability or information to properly assess computer-generated suggestions. Too much trust in the results provided by the computer leads to an 'automation bias'.

[194] Paul de Hert, Hans Lammerant, "Predictive profiling and its legal limits: Effectiveness gone forever?", in: Bart van der Sloot et al. (Eds.) "Exploring the Boundaries of Big Data"

5. Implications of the GDPR - A media analysis

In this section, we present a short summary of the results of a media analysis on GDPR-related coverage. Media analysis is a sub-set of content analysis that applies a systematic method to study mass media as texts, including the content of newspapers. The analysis focused on two countries, namely Germany and the UK, and examined a sample of eight different sources dating from 15 May to 15 July 2018, with an emphasis on frequently appearing issues, the portrayal of consequences of the GDPR and references to technical solutions. We provide a detailed description of the theoretical concepts, the research approach, the content identification and analysis method, and the results in the appendix.

The analysis shows that in both countries the amount of articles that deal with the GDPR is comparable and that the intensity of reporting follows the same pattern; in both countries the peak of reporting was around 25 May (the date when the GDPR became effective). Already four weeks after the beginning of the monitoring there was hardly any coverage at all.

To investigate the covered issues within GDPR-related coverage, we focused on three classes of categories: (1) general issues that appeared frequently, (2) consequences of the GDPR and (3) references to technical solutions in support of GDPR compliance. Regarding the general issues, we examined the following issues:

- *General information on the GDPR*
  We found that roughly one third (32.3%) of the German and one fourth (22.4%) of the UK sample contain general information on the GDPR. This was coded when an article dealt with when the regulation would become effective, who is affected and what might happen if someone is not compliant. These articles were found most frequently around 25 May.
- *Difficulties for small businesses*
  This discourse was found to be far more prominent in German than in British coverage (28.2% vs. 13.3%). It deals with difficulties that small businesses and other organisations like local authorities, associations, bloggers face due to the new regulation, as it is based on the idea that while large companies will normally have the resources to prepare for the GDPR to an appropriate extent, this will be hard for small companies that may not have the funds for extensive legal advice.
- *The role of Facebook & Co.*
  This reference was found in 39.5% of the German and 33.6% of the UK sample, mainly deriving from a company being found to be privacy intrusive. Facebook often seems to stand for the US technology giants. Other companies such as Apple, Google and Microsoft are also mentioned, although to a lesser extent.
- *The Cambridge Analytica scandal*
  The Cambridge Analytica data scandal was coded as a separate category, as it appeared in roughly half of the coverage referencing Facebook and thus in 19.4% of the German and 16.1% of the British sample. In short, the personal data of tens of millions of people harvested by Facebook was shared with the political consultancy Cambridge Analytica from 2014 on and used to influence voter opinion.
- *Data as the currency of the 21st century*

This discourse appears in 11.3% of the German and 10.5% of the UK coverage and deals with the increasing importance of data flows, with data brokerage being one of the fastest growing sectors of the economy. In this regard, data has been described as the "new oil" of the global economy.

Especially in the very first days of our period of research and also from the middle of June on, there are many articles that only refer loosely to the GDPR. The articles have main topics other than the GDPR. This reflects the declining interest in the GDPR after the new regulation became effective on 25 May. Articles assigned to this category make up 31.5% of the German and 47.6% of the UK sample.

With respect to consequences of the GDPR, we examined the following discourses:

- *Restrictions on the provision of services*
  From the articles that contain information on consequences of the GDPR, 20.3% in the German and 26.3% in the UK sample refer to the limitation of access of EU users to certain international services, including such prominent examples as the LA Times. While most websites, according to the articles in the sample, stressed that the cutoffs were temporarily while implementing the required technical compliance solutions, many of them are still not available to EU readers (as of 7 October 2018).
- *A wave of cease-and-desist letters*
  While this discourse is very prominent in the German reporting that deals with the consequences of the GDPR (37.5%), it is much less frequently found in the respective British coverage (15.8%). This was because political efforts by German political parties to prepare for a wave of cease-and-desist letters were only found in the German media.
- *Countless re-subscribe e-mails*
  In the weeks leading to the enforcement of the GDPR, consumers were inundated with e-mails asking them to "opt in" to continue receiving material. This issue is covered quite extensively in both the German and the British media, as it is referred to in 29.7% of the German and even 64.9% of the British articles that deal with consequences of the GDPR, constituting the by far most frequently mentioned consequence of the GDPR in the British sample.
- *Ban of social media on business smartphones*
  Automotive giant Continental has banned workers from using WhatsApp and Snapchat on their smartphones over concerns the apps might be leaking confidential information to third parties. While this issue seemed to be prominent during the exploratory work leading up to the development of the categories, it actually is mentioned in only 6.3% of the German articles that deal with consequences of the GDPR and even only 1.8% of the respective British ones and thus does not build a prominent part of the GDPR-related discourse in both countries.
- *European competitiveness*
  An issue that is of special interest to our research is the question if the media would portray possible consequences of the GDPR with regard to the future role of Europe in the international context. Our research shows that indeed there was such a discourse, as it was found in 31.3% of the German articles that deal with the consequences of the GDPR and 22.8% of the respective British ones. While in the German media the assessment was found to be quite balanced with both positive and negative consequences being mentioned in 40% of the articles, in the respective British articles, only 15.4% mention positive consequences for the EU while 76.9% focus on the negative side.

Regarding the references to the need for technologies to act in a GDPR-compliant way, we used the following classes of technologies: anonymisation, sanitisation, encryption, deletion, multi-party computation, access control, policy enforcement, accountability, data provenance, transparency, access and portability, and user control. In both countries, there were no references to anonymisation, multi-party computation and data provenance, whereas in both Germany and the UK references to deletion, transparency, access and data portability and user control were found frequently. References to encryption, and accountability and audit mechanisms were found less frequently.

Regarding possible next steps, there are several aspects that would be interesting to investigate further. On the one hand, it would be interesting to take a deeper look into the respective media's portrayal of certain aspects. While the described analysis focused on comparing two countries, it was visible for instance in the discourse "future role of Europe" that there were significant differences in the portrayal of the issues. There were indications that the conservative press tended to portray the consequences more negatively. It is likely that the respective sources differ also with respect to other aspects. However, to investigate this properly, it would be helpful to analyse a bigger sample to ensure the analysis is still valid. This could, for instance, be reached by extending the period of analysis or the keywords to source relevant articles.

On the other hand, it would also be interesting to compare other countries. The US would be an interesting choice in this regard, given the fact that most of the giant technology companies are based there and the role that regional and cultural differences play with respect to the availability and use of privacy-preserving technologies.[195]

[195] See D3.2 of eSIDES.

6. Conclusions

This Deliverable provides a gap analysis related to the findings of Deliverable 2.2 and Deliverable 3.2 of the e-SIDES project. Deliverable 2.2 provides a general assessment of ethical, legal, societal and economic issues that emerge in different big data contexts in general, while Deliverable 3.2 provides a technology-specific assessment of classes of currently existing privacy-preserving technologies including their effectiveness in addressing ethical and societal issues and the challenges that arise in their implementation. In Deliverable 3.2 we concluded that ethical and societal issues remain present, mainly because available technologies are not widely integrated into big data solutions. This implementation gap is the focus of this report.

Two broad categories of gaps can be distinguished. First, there may be issues for which solutions are either insufficient or do not exist. Second, there may be issues for which solutions do exist, but those solutions are not being implemented. The first type of gaps could be addressed by further developing design requirements for privacy preserving big data technologies and innovating new types of solutions. For the second type of gaps, there is not so much need for further technological developments, but rather for finding the reasons for which the existing and available solutions are not used and implemented. Perhaps the solutions are unknown or too expensive or not prioritized. Depending on the reasons behind this, next steps can be formulated to address the ethical, legal, societal and economic issues.

Throughout this deliverable we maintain that the ethical and legal reasons for the implementation gap are closely intertwined with the social and economic reasons. We have focused on them separately in order to highlight the challenges and opportunities from each of these perspectives in closing the gap.

In the section on legal and ethical aspects, we highlighted that the explanations for the limited or scarce introduction of privacy-preserving technologies lies to a large extent in their flexible interpretation being both a blessing and a curse for practitioners. This point has also been strengthened by the economic and social sub-section where the refrained introduction of privacy-preserving big data technologies was aligned with organisational struggles in defining privacy too narrowly, conflicting definitions and troubles in translating such definitions into design requirements.

With respect to further legal and ethical reasons specifically in the context of sensitivity of data, we found that although the GDPR offers clear benefits for protecting a broad set of sensitive information, not all professionals embrace the potential for strengthening the right to non-discrimination to the same extent, which right is clearly facilitated by Art. 9 of the GDPR. For instance, healthcare researchers argue that strict privacy protection and also the robust implementation of privacy-preserving technologies could hamper epidemiology research and big data's value for the advancement of healthcare research.

A strong component of the ethical and legal reasons leading up to or stemming from the limited or scarce introduction of privacy-preserving technologies relates to a number of liability and ethical responsibility considerations that need consideration. Organisations that collect, use and distribute data are in general responsible for privacy-related tasks such as anonymisation and encryption. Data controllers and processors have both legal and ethical obligations to implement such privacy-preserving technologies. When viewing the entire data value chain, privacy preservation must be a shared responsibility. Yet, in effect the levelling of responsibilities should correspond with the strength of a given party involved in the data value chain. This is currently not the case. Yet, an overall transparency for the data value chain is

crucial in respect to the liability and responsibility of stakeholders so that they can be taken into account for the development and implementation of big data solutions. Currently, regional differences in data protection and antitrust or competition law regimes between the EU and the US also influences the extent to which liabilities of big data stakeholders are made transparent and the extent to which privacy-preserving technologies are embraced and implemented. Ideally the legal regimes of data protection, which should help to stem data breaches, and competition law, which should help prevent certain organisations from utilising a dominant position in the market, in general should complement and desirably mutually strengthen each other[196]. This (the enforcement of data protection and competition law) is indispensable to both protect the rights of consumers and the rights of SMEs in doing big data-based business. On this subject US and EU approaches differ, yet some movements of the US antitrust regulators from September 2018 begin to point towards the EU's approach (for more, see footnote 91 in this document). This change in the US approach includes not only that criteria to measure price discrimination should be leading but also a general vision change for companies to strive for a more equal treatment of consumers. In that regard, privacy-preserving technologies could play a facilitating role. The European Union's data protection supervisor, Giovanni Buttarelli, envisions probably the most fruitful approach: he sees collaboration between the different regulatory regimes of data protection, consumer protection, and antitrust in order to address the platform power of organisations and its impact on consumers.[197]

As much as these legal and ethical reasons and challenges are intertwined with each other so are the economic and social reasons that explain a lack, or limitations of incentives to introduce privacy-preserving technologies. One of our findings in the economic and social reasons sections relates first to the fact that, for instance, cutting-edge hardware and software is likely to be more expensive than solutions with a track-record. Although costs of resources needed in order to respond effectively are relevant in the privacy context, the costs related to the resources instrumental to repair damaged systems and data are often beyond the calculation of companies. This closely ties into our third finding, that from an economic point of view, putting the foundational principles of data protection into practice can be in conflict with business models. Yet with the GDPR going into effect in May 2018, some companies might have accepted the risk of ignoring data protection compliance requirements. In this sense the GDPR can help; especially, given the extensive fines, which are now more likely to follow from non-compliance, more and more companies are forced to adopt substantial changes. Our fourth finding with respect to the economic and social aspects was that business model conflicts often result from trade-offs drawn between privacy preservation and data use. One area of big data context where such business model is used is healthcare, and as explained above healthcare researchers have particular reservations as to introducing too strict privacy-preserving technologies as they fear that could impede innovation. Yet, our fifth finding underlines that the privacy, security and data protection standards drawn up in Europe have the capacity to become a trademark even beyond the EU. The developments in the legal framework facilitate a movement into this direction, which points towards addressing key economic and social reasons for the implementation gap.

[196] Natasha Lomas, "Europe is drawing fresh battle lines around the ethics of big data", TechCrunch (3 October 2018), https://techcrunch.com/2018/10/03/europe-is-drawing-fresh-battle-lines-around-the-ethics-of-big-data/.
[197] Ibid.

With respect to economic and social aspects in embracing privacy-preserving technologies a sixth point to make relates to cultural differences. Bridging cultural differences is challenged by the fact that privacy outcomes are often unpredictable and privacy concerns and expectations are context-dependent. For instance, cultural values can influence people's privacy perceptions such that countries with tighter privacy regulations experience fewer privacy problems.[198] Yet, as our media analysis has also shown there were significant differences in the portrayal of the issues around the implementation of GDPR, including privacy-preserving technologies, if comparing the coverage in Germany and the UK. A large portion of cultural reasons, we argue, however could be addressed by offering individuals personal benefits – not only free access to online services, but monetary compensation – for the value of their data being used by big data companies. A seventh economic and social reason to mention is that despite data breach scandals and the rise of cyberattacks, consumers are still often attracted to buy products and services from a provider who offers lower prices and lower data protection than companies offering a higher level of data security and higher prices for products and services. This requires on the one hand a consumer mentality change, but also the acquisition of new skills for all stakeholders in the value chain. Developers need knowledge not only of maths and statistics, but also insight into data ethics, information law, and privacy law, among others. Users need to be tech savvy, aware of privacy risks, and have the skills to prevent or combat such risks. Furthermore, the amendment of business models towards a more value-sensitive approach could complement the process and motivate the embracement of privacy-preserving technologies. The capacities of the new EU data protection and competition law can clearly promote the integration of privacy-preserving technologies into big data solutions. Furthermore, the role of technologies in closing the implementation gap is also crucial to embrace. An increase in proactive approaches by privacy-preserving technologies that prevent breaches or rule violations in the first place are stimulated by the legal regime but shall desirably also be embraced as a mentality change. Furthermore, smaller companies might want to initiate competition law cases (as a violation of the freedom to do business) towards larger companies as well, especially after Google's antitrust case regarding Android in the EU.[199]

As a final concluding remark for our analysis of ethical legal, economic and societal aspects leading up to and stemming from the limited or scarce introduction of privacy-preserving technologies we also stress the need for the constant reassessment of privacy-preserving technologies with respect to their design and use in relation to the social, ethical, legal and economic effects on persons and society at large. As we have demonstrated, these technologies alone cannot address all ethical and societal issues. A combination of legal, ethical, economic and societal changes are needed in order to facilitate transformations towards a more privacy-preserving data usage that capitalizes on the value of big data but respects the ethical, legal and societal limitations of exploiting that data. Privacy-preserving technologies can facilitate transformations in business models that are mainly aimed at data exploitation, in organisational structures and towards more cultural receptiveness of these technologies after walls around differences in definitions and expectations of privacy preservation are cracked down, or made more transparent

[198] Sara Dolnicar, Yolanda Jordaan, "A Market-Oriented Approach to Responsibly Managing Information Privacy Concerns in Direct Marketing", doi:10.2753/JOA0091-3367360209

[199] Jennifer Rankin, "Google fined £3.8bn by EU over Android antitrust violations", The Guardian (18 July 2018), https://www.theguardian.com/business/2018/jul/18/google-faces-record-multibillion-fine-from-eu-over-android.

among stakeholders. Each of these steps should bring us closer to a wider implementation of privacy-preserving technology solutions in the variety of big data contexts.

Appendix

This appendix describes the media analysis that led to the results summarised in section 5.

In the EU, from May 2018 onwards, all processing of personal data has to be made in observance of the legal provisions of the EU GDPR. The addressees of this EU regulation are required to meet certain objectives at their own discretion and to make their own business decisions in order to comply with the following principles:

- Lawfulness, fairness and transparency;
- Purpose limitation;
- Data minimisation;
- Accuracy;
- Storage limitation;
- Integrity and confidentiality.

Therefore, all big data projects have been or should be assessed bearing all these legal principles in mind.[200] Under the GDPR, organisations can face sanctions of up to 4% of their annual gross global revenue in the event of a breach or data mismanagement.[201]

The GDPR applies when a data controller or a data processor is based in the EU, but it also addresses the transfer of data outside the EU if the data collected or processed is from individuals based in the EU. Personal data in this regard refers to any information connected to an individual, but there are exceptions for data processed in an employment context or for national security purposes that are subject to national regulations. To be compliant with the GDPR, data controllers need to design data protection measures into their business processes for products and services.[202]

Our aim is to conduct a media analysis that is supposed to show how the GDPR is reflected in the media with an emphasis on the consequences for companies. What have (multinational) companies done when the GDPR became effective? Which strategies are reflected in the media? Which discourses revolve around them? Are there any reliable assessments of the long-term effects for Europe and European companies? How frequent are respective articles in the different media? Is there still reporting after the GDPR has become effective? Research on these issues can help us gain a better understanding of the regulatory, organisational and research gaps and how they are perceived.

Special attention is paid to technical solutions that help companies and service providers to act in a GDPR compliant way. Are these solutions mentioned at all in GDPR-related coverage? If so, we are interested if these solutions have the potential to contribute to the gap analysis as described in sections 2 and 3. In this regard, we refer to Article 25 of the GDPR on data protection and design by default, which says:

1) *Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and*

[200] Rafael García del Poyo, Roger Segarra and Samuel Martínez "Big Data analysis and anonymisation techniques under the EU General Data Protection Regulation**",** https://www.financierworldwide.com/big-data-analysis-and-anonymisation-techniques-under-the-eu-general-data-protection-regulation/#.Wqem12f3i1l (accessed 8 May 2018)

[201] Mekhala Roy, "Data anonymization techniques less reliable in era of big data"

[202] EU GDPR.ORG, https://www.eugdpr.org/

freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

2) *The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.*
3) *An approved certification mechanism pursuant to Article 42 may be used as an element to demonstrate compliance with the requirements set out in paragraphs 1 and 2 of this Article.*[203]

We study if the mentioned technical and organisational measures appear in GDPR related coverage and if so, which do appear most often.

The structure of this appendix is as follows: First of all, we outline important theoretic concepts that frame our research. Afterwards, we give a brief introduction to our research approach, describing the content analysis process and the selection of the sample. Finally, the results are presented with an emphasis on the issues most commonly found in GDPR-related coverage, the consequences that are addressed and the potential gaps that occur.

Theoretical concepts and research approach

In the information society, the public's perception of privacy and its ethical, legal, societal and economic implications are reflected and reinforced by the media. The media's role in reconstructing images, perceptions and beliefs is crucial; media express and at the same time shape public opinion. Prior research has shown that generally there are strong interrelations between public discourses in media reporting and the individual as well as collective perception of the covered issues. The common methodology to investigate the media reporting on a certain issue is to conduct a systematic content analysis. According to Krippendorff, "content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use."[204] Content analysis is a methodology that can be applied to a huge amount of texts by dividing them into the key features of interest that are coded during the coding process. Although there are a variety of approaches such as (critical) discourse

[203] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)", *Official Journal of the European Union*, L 119, 04 May 2016, pp. 1-88. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=DE

[204] Krippendorff, Klaus, *Content analysis: An introduction to its methodology*, 3rd ed. (Los Angeles, London: SAGE, 2013), p. 24.

analysis[205], social constructivist analysis[206], rhetorical analysis[207], ethnographic content analysis[208] and conversation analysis[209], the approaches share some common characteristics. Sometimes referred to as interpretive, all of them require a close and detailed reading and involve the rearticulation into new narratives and thereby interpretation of the analysed texts in regards to critical, deconstructive, emancipatory and other scholarly traditions. Given the fact that unlimited resources in scientific research are the absolute exception, it becomes obvious that only a relatively small sample can be analysed qualitatively in the above described sense.[210]

In line with Krippendorff, Früh and Mayring, we do not regard the harsh distinction between quantitative and qualitative approaches as helpful for our research. Every quantitative analysis includes several qualitative steps, such as exploring and reading (parts of) the material, the development of the coding guidelines and finally the interpretation of the results – not to speak of the fact that every research question derives from, if one wishes to use the terminology further on, observations and problems that are regarded as qualitative. In addition, content analysis is never only about quantities as a statement of fact, but about what these quantities mean in a broader sense. Also the other way round most so-called qualitative analyses base their results on some quantification of the analysed cases, such as the frequency and intensity of a specific issue mentioned in a single or several texts or interviews.[211] Nevertheless, there are obviously more quantitative and more qualitative aspects of research, which can be fruitfully combined in different phases of the research process. In this regard, we follow Weber, who sees the combination of both quantitative and qualitative approaches as the best way to design content analyses in a meaningful way.[212]

'Meaningful' in this regard means being able to draw inferences from the analysed texts to a specific context. Inferences are conclusions that the researcher draws from the analysed texts to a context. This context has to be chosen by the researcher and may include, among others, the speaker/producer of the analysed texts, the reader who is supposed to read those texts, and the historical, political or social situation.[213]

Content identification and analysis

Ever since the terrorist attacks in September 2001 on the US, the media have built a strong narrative around the need to give up privacy for more security. This narrative was intensified after the Boston

[205] van Dijk, T. A., "Principles of Critical Discourse Analysis," *Discourse & Society* 4, no. 2 (1993): pp. 249–83.
[206] Gergen, Kenneth J., "Social constructionist inquiry: Context and implications," in *The social construction of the person*, eds. Kenneth J. Gergen and Keith E. Davis (New York: Springer-Verlag, 1985).
[207] Harris, Karen L., "Content analysis in negotiation research: A review and guide," *Behavior Research Methods, Instruments, & Computers* 28, no. 3 (1996): pp. 458–67.
[208] Altheide, David L., "Reflections: Ethnographic content analysis," *Qualitative Sociology* 10, no. 1 (1987): pp. 65–77.
[209] Sacks, Harvey, "An analysis of the course of a joke's telling in conversation," in *Explorations in the ethnography of speaking*, eds. Richard Bauman and Joel Sherzer (London, New York: Cambridge University Press, 1974).
[210] Krippendorff, 2013, pp. 22–23.
[211] Ibid.; Früh, Werner, *Inhaltsanalyse: Theorie und Praxis*, 6., überarbeitete Auflage (Konstanz: UVK-Verl.-Ges, 2007); Mayring, Philipp, *Qualitative Inhaltsanalyse: Grundlagen und Techniken*, 11., aktualisierte u. überarbeitete (Weinheim: Beltz, 2010).
[212] Weber, Robert Philip, *Basic content analysis*, 2nd ed. (Newbury Park, Calif: Sage Publications, 1990).
[213] Rössler, Patrick, *Inhaltsanalyse*, 2., überarbeitete Auflage (Stuttgart: UVK-Verl.-Ges., 2010), pp. 31-36, 237-240.

attacks in April 2013.[214] At the same time, it is known that the media are not neutral and independent accumulators and transmitters of information, but rather produce a certain kind of reality via selection and interpretation. The theory of news values emphasises that news narratives which contain an amount of, among others, negativity, sensationalism, simplicity and proximity are more likely to be published.[215] In this regard, the revelations by Edward Snowden about the mass interception programmes from June 2013 mark a disruptive event that had enormous consequences for the media's portrayal of privacy and security-related issues.

Here, we are interested in the media reporting on the GDPR in Europe. Due to time and budget constraints and with regard to availability and language capabilities, we focus on two countries; namely Germany and the UK. This choice is especially interesting with regard to privacy-preserving technologies as previous research has found that media discourse revolving around privacy and security-related issues is framed quite differently within the two countries: In short, discourse in Germany was found to be privacy-centred, whereas in the UK it is security-centred. Elsewhere[216], we have analysed media coverage before and after the Snowden revelations in both of the countries (as they had contradictory settings in the affair) and found that from 2008 to 2011, privacy and security-related discourse in both Germany and the UK is to a vast extent concerned with issues revolving around data and personal information, encompassing at least partly a so-called 'warning-narrative', as it refers to events that are described as threating privacy in the sense of personal data and information and the security of these. Regarding the notions of privacy and security, although there are variations, there was a tendency of privacy being used most frequently in the sense of privacy of personal data, while security was used more ambiguously.

What the countries differ in is whether the one or the other is in the foreground; accordingly, the discourse is considered either privacy or security-centred. In fact, the discourse in Germany showed to be clearly privacy-centred, with an emphasis on the risks that the concept is subject to and a strong focus on the need for protection. In the UK, the discourse is security-centred. The additional sample that represents discourse directly after Snowden's revelations in 2013 reveals that these tendencies stay stable in both of the countries: While in Germany, the focus is still on the amount of privacy intrusion that became known with the revelations, in the UK the focus on security – and to be precise, on national security – is even reinforced; although there are variations between the different sources. Issue-wise, in both countries the privacy and security-related discourse is completely reflected in the sense of interception and spying; the affair and its consequences are clearly dominating. While in Germany, the different sources are quite

[214] Visible for example in Time magazine's 2013 may issue "Do we need to sacrifice privacy to be safer?"; see Calabresi 2013.

[215] See for example Galtung, J. and M. H. Ruge, "The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers," *Journal of Peace Research* 2, no. 1 (1965): pp. 64–90.; Schulz, Winfried, *Die Konstruktion von Realität in den Nachrichtenmedien: Eine Analyse der aktuellen Berichterstattung*, 1st ed. (Freiburg [Breisgau], München: Alber, 1976); Staab, Joachim Friedrich, *Nachrichtenwert-Theorie: Formale Struktur und empirischer Gehalt* (Freiburg: K. Alber, 1990f).

[216] Weitkamp, Jana, Simone Kimpeler, and Michael Friedewald, "Content and discourse analysis of security and privacy reporting in the European media", PRISMS Deliverable 6.2, April 2015. http://publica.fraunhofer.de/documents/N-336428.html

homogenous in their evaluation of the affair, the British sources differ; especially The Guardian and The Daily Telegraph.[217]

Within the two chosen countries, we looked for two main national, daily, quality newspapers and one tabloid/popular newspaper. We base our choice to analyse printed press on findings that the usage of newspapers mainly derives from the wish to find relevant information, and on the assumption that their effect in terms of remembering what was read is higher in comparison to broadcast media, as a more active examination of the material is needed.[218] In addition, although a wide variety of alternatives to the traditional news media exists (e.g., blogs), conventional mass media still has an important societal relevance by enabling public communication. The increasing amount of ever more news providers and platforms that focus on a special topic at the same time increase the perceived importance of traditional mass media, as they only present a selection of topics and thus provide a focussed view on issues that are deemed relevant.[219] Thereby, they offer orientation especially when it comes to complex topics. Furthermore, newspaper articles are relatively well archived, so that the availability and accessibility is much higher. This is especially important regarding the above expressed claim that content analyses have to be replicable. Nevertheless, one cannot deny that the importance of online media is increasing, so we are also analysing one (major) news website per country.

Criteria for our choice of newspapers were national availability, circulation and political orientation; which means that we sourced the two most important newspapers with different political stance. To find relevant online news websites, we additionally considered access rates per day and availability. National quality newspapers often function as (opinion) leading media, which are not only influential to political elites and thereby have high impact on what gets into the political and public agenda, but that are also widely read by other journalists and thus published by other media.[220] Thus, we assume that the reported issues in our sample are far more widely distributed than only in the media we analyse and thus read by more people. Table 1 shows the selected media.

	Liberal	**Conservative**	**Tabloid**	**Online**
Germany	Süddeutsche Zeitung (SZ)	Frankfurter Allgemeine Zeitung (FAZ)	BILD	Spiegel Online
UK	The Guardian	The Daily Telegraph	Daily Mail	BBC News

Table 1: Overview of selected media

As we are especially interested in the consequences of the GDPR, the analysis was carried out as an ongoing monitoring from the GDPR becoming effective minus roughly a week and the following two months and therefore dates from 15 May until 15 July 2018. To find relevant articles, we used a

[217] Ibidim.

[218] Linzmaier, Vera, Lebensmittelskandale in den Medien: Risikoprofile und Verbraucherverunsicherung (München: Reinhard Fischer, 2007), p. 117.

[219] Jarren, Ottfried, "Massenmedien als Intermediäre: Zur anhaltenden Relevanz der Massenmedien für die öffentliche Kommunikation," *M&K* 3-4 (2008): pp. 329–46, http://www.m-und-k.nomos.de/fileadmin/muk/doc/Aufsatz_Muk_08_3-4.pdf (accessed 15 March 2013).

[220] Gerhards, Jürgen and Mike Steffen Schäfer, *Die Herstellung einer öffentlichen Hegemonie: Humangenomforschung in der deutschen und der US-amerikanischen Presse*, 1st ed. (Wiesbaden: VS Verlag für Sozialwissenschaften, 2006), p. 74; Neidhardt, 1994.

straightforward approach by using the key term *GDPR OR "general data protection regulation"* in English and *DSGVO OR Datenschutzgrundverordnung* in German as tests have shown that all articles dealing with the issue in fact use either the abbreviation or the complete term. The material for the analysis was extracted using a keyword search from a specialised full-text database (LexisNexis) for all sources except BBC News, which we sourced directly from its website.

To investigate GDPR-related coverage, our approach was twofold: On the one hand, we gained a rough overview of our sample by creating a tag cloud of the most prominent words. On the other hand, a systematic random sample within our sample was explored as a basis to gain an understanding of the discourse and covered issues and to develop a set of categories for coding. During the coding process, this system of categories was revised if other issues of interest appeared.

Before exploring the sample, we had a few assumptions that led to the first set of categories: We expected to find a lot of articles that would directly deal with the GDPR and what it means for different groups including consumers and companies. We also expected to find many references to *Facebook* and the *Cambridge Analytica* incident. We also expected to find articles dealing with the consequences of the GDPR, especially with the cancellation of services of some US platforms and the ban of social media apps on company smartphones. These assumptions were confirmed by the tag clouds and the exploration of the random sample. However, during the exploratory phase we also found other issues of interest, which resulted in the following set of categories:

- Firstly, general issues that seemed to appear frequently like general information on the GDPR and its coming into force and the difficulties it brings for organisations to be compliant. Other issues within this class of categories include coverage related to Facebook and other technology companies in general and the Cambridge Analytica scandal in particular, a discourse that we named "data as the currency of the 21st century" and lastly various articles that only loosely refer to the GDPR but do not foreground it.
- Secondly, we put emphasis on the consequences of the GDPR that are shown by the media. Covered issues within this set include the cancellation of services of some US platforms, the fear of fines, countless re-subscribe e-mails being sent in a struggle to become GDPR compliant and the ban of social media apps on company smartphones. We were especially interested if the current and future role of Europe in an international context would show up in our sample and if so, if the consequences of the GDPR are portrayed more positively or negatively.
- Lastly, we paid special attention to technical measures. It was coded if these are mentioned in GDPR-related coverage and if so, it was investigated if it was possible to categorise these within the project's classes of technologies.[221]

Results

Table 2 and Table 3 provide an overview of our sample by listing the sources and the numbers of articles found. To get a first impression of the respective country's and media's sample size, the total number of words and the average length per article are also shown.

[221] See Bachlechner, Daniel, Michael Friedewald, Jana Weitkamp, and Nicholas Martin, "Overview of existing technologies", e-SIDES Deliverable 3.1 e-Sides Project, 2018. https://e-sides.eu/assets/media/e-SIDES%20D3.1%20V1.0.pdf

	Liberal: Süddeutsche Zeitung	**Conservative: Frankfurter Allgemeine Zeitung**	**Tabloid: BILD**	**Online: Spiegel Online**	***Total***
No. of articles	49	47	4	24	124
No. of words (total)	29,978	42,276	1,537	17,310	91,101
Average words per article	611.8	899.5	384.3	721.3	734.7

Table 2: Sample overview Germany

	Liberal: The Guardian	**Conservative: The Daily Telegraph**	**Tabloid: Daily Mail**	**Online: BBC News**	***Total***
No. of articles	47	49	16	31	143
No. of words (total)	75,445	23,961	8,352	18,004	125,762
Average words per article	1,605.2	489.0	522.0	580.8	879.5

Table 3: Sample overview UK

Table 2 and Table 3 show that the samples gained are quite comparable in absolute numbers. Most articles on the GDPR were found in the two quality newspapers, both in Germany and the UK. To a lesser extent, coverage took place in the online sources. Only few articles were found in the popular newspapers. Whereas the respective sources differ significantly, the average length per article is also comparable in the two countries with 734.7 words in Germany and 879.5 words in the UK. In the UK, one source stands out with extensive and lengthy coverage: The Guardian's average length per article is 1,605.2 words.

Intensity of reporting

Before presenting the core themes found in the sample, we start by having a general look at the nature of our data. First of all, we were interested in how the intensity of reporting evolved over time. The development of reporting within the selected time span is shown in Figure 5 and Figure 6.

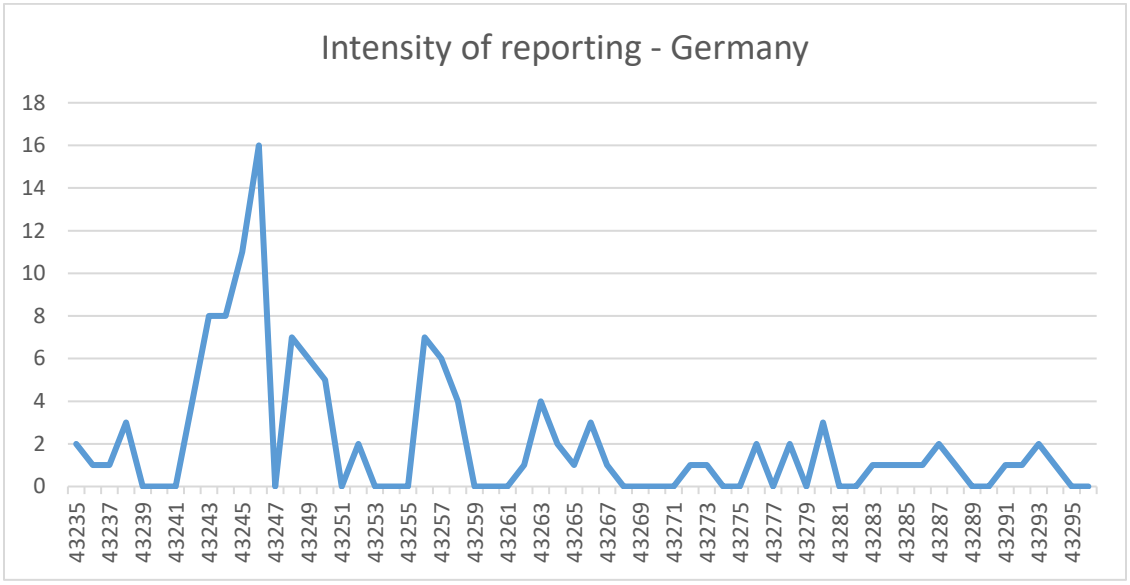

Figure 5: Intensity of reporting between 15 May and 15 July 2018 in Germany

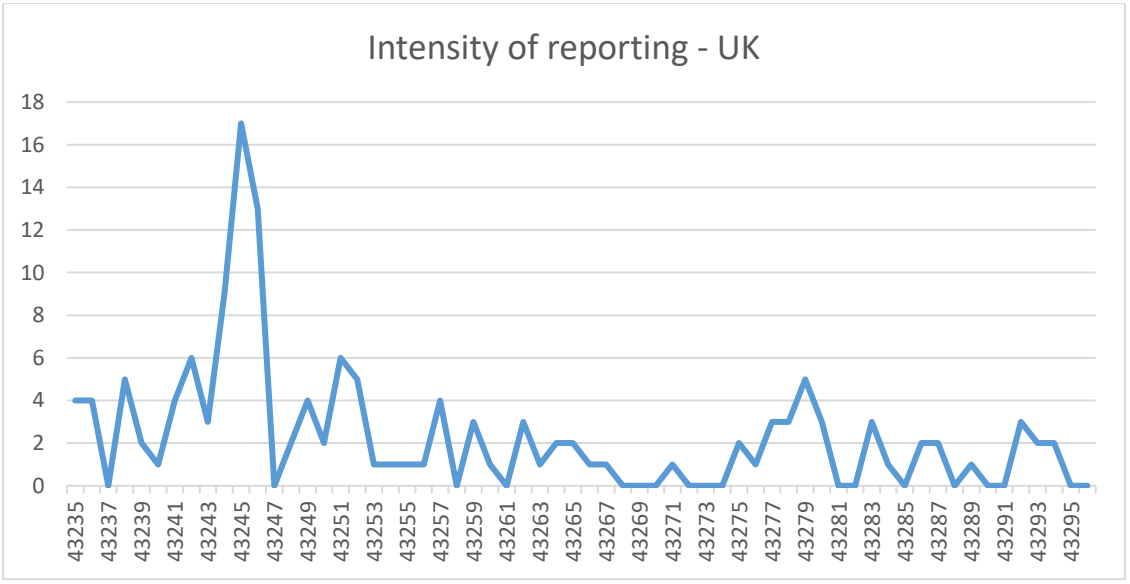

Figure 6: Intensity of reporting between 15 May and 15 July 2018 in the UK

The intensity of coverage had its peak around the date when the GDPR became effective on 25 May 2018 with most of the articles being published between a few days before and a few days after that particular date. Already four weeks after the beginning of the monitoring, there is hardly any coverage at all. The investigation of the covered issues shows that during this peak of reporting, most of the articles deal directly with the GDPR becoming enforceable and providing general information, whereas from the middle of June until the middle of July, the GDPR is mostly only referred to, but not the main topic of the published article. This comes as no surprise as the influence of events on the media's reporting has always been crucial.

Authors

Looking at the authors of the analysed articles, Figure 7 and Figure 8 show that the vast majority of GDPR-related coverage in both Germany and the UK stem from in-house journalists (83.9% in Germany and 74.8% in the UK). There are hardly any articles at all that derive solely or partly from news agencies (0.8% in Germany and 0% in the UK). However, the portion of articles in the UK sample where the authors are not identifiable is rather high (15.4%). This is mainly deriving from the BBC News coverage, where authors are only mentioned occasionally. Another thing to note is that the portion 'Guest author' is bigger in Germany than in the UK (6.5% vs. 4.2%), although the portion 'Letters to the editor' is smaller (1.6% in Germany vs. 5.6% in the UK), indicating that articles by authors, who are not part of the respective newspaper/online source, are published more regularly in Germany, whereas in the UK, it is more the 'reader's voice'.

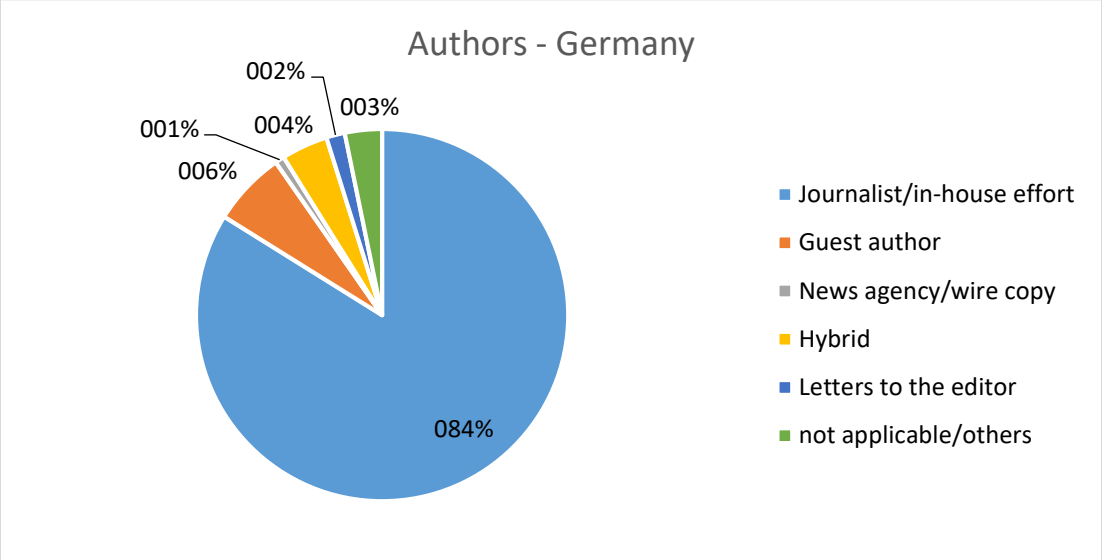

Figure 7: Authors in the German sample

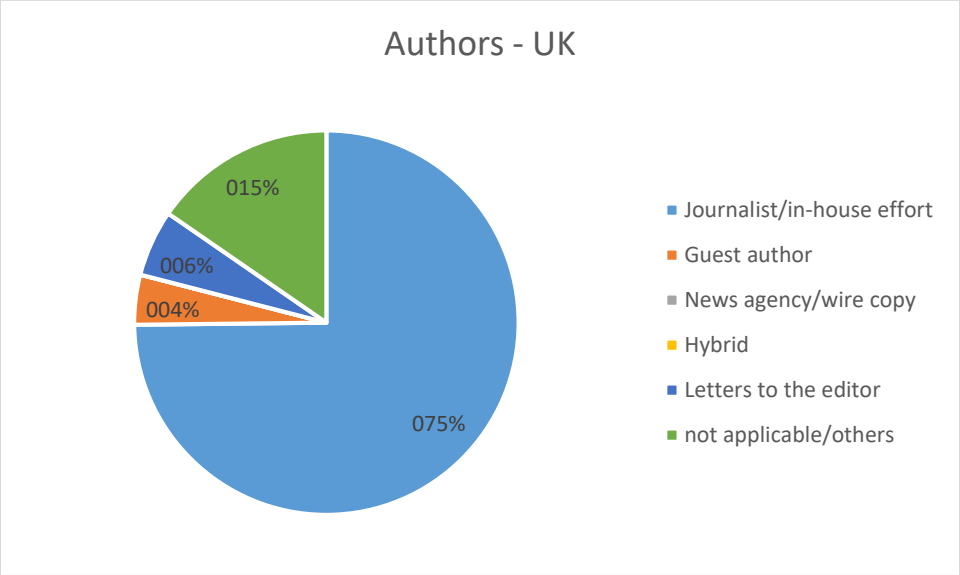

Figure 8: Authors in the UK sample

Sections

The examination of the sections, which is illustrated in Figure 9 and Figure 10, shows that GDPR-related coverage in Germany takes place most frequently in the Economics and Business section (40.3%), followed by the sections Technology, Computer & Internet (16.1%), Local News (12.1%) and Politics (10.5%). In the UK, the distribution is different: GDPR-related coverage takes place most frequently in the section Technology, Computer & Internet (32.2%) followed by the sections Economy and Business (28.0%) and Politics (10.5%). However, the share of articles where the section could not be identified is also rather high in the UK sample (10.5%), which derives mainly from the Daily Mail's coverage, where information on sections is not given in the database. As it was already indicated, the share of 'Letters to the editor/opinion' is also rather high in the UK sample (9.8%).

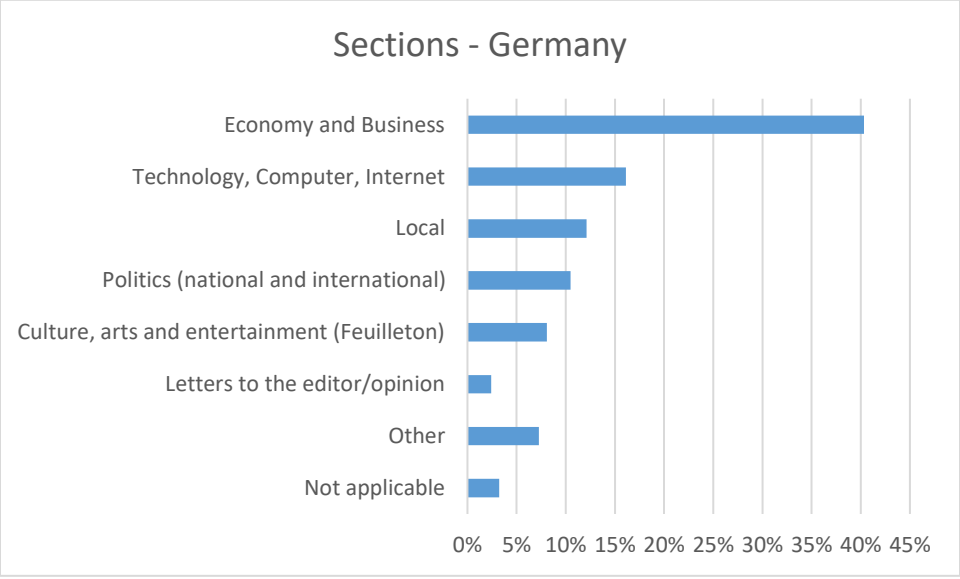

Figure 9: Sections in the German sample

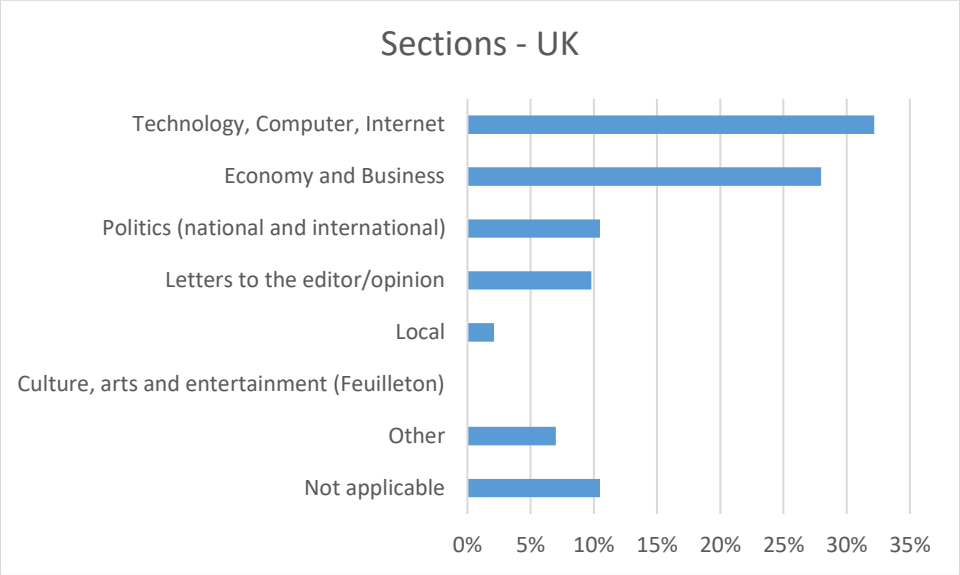

Figure 10: Sections in the UK sample

Covered issues

Our approach to investigate GDPR-related coverage was twofold: On the one hand, a tag cloud of the most prominent words was generated to gain a first impression of the GDPR-related discourse in the two countries. On the other hand, a systematic random sample within our sample was explored as a basis to gain an understanding of the discourse and covered issues, and to develop a set of categories for coding. During the coding process this system of categories was revised if other issues of interest appeared.

Figure 11 Most frequent words used in the German GDPR-related coverage

Figure 12: Most frequent words used in the UK GDPR-related coverage

The tag clouds generated on the basis of our plain text sample (see Figure 11 and Figure 12) give a good impression of what is covered frequently in the media's portrayal of the GDPR. They show the 50 most frequently appearing words in each sample and thereby represent each country's discourse.

Combined with the exploratory work, we ended up coding three classes of categories: Firstly, general issues that seemed to appear frequently like general information on the GDPR and its coming into force and the difficulties it brings for organisations to be compliant. Other issues within this class of categories include coverage related to Facebook and other technology companies in general and the Cambridge Analytica scandal in particular, a discourse that we named "data as the currency of the 21st century" and lastly various articles that only loosely refer to the GDPR.

Secondly, we put emphasis on the consequences of the GDPR that are shown by the media. Covered issues within this set include the cancellation of services of some US platforms, the fear of fines, e-mails being sent in a struggle to become GDPR compliant and the ban of social media apps on company smartphones. We were especially interested if the current and future role of Europe in an international context would show up in our sample and if so, if the consequences of the GDPR will be portrayed more positively or negatively.

Lastly, we paid special attention to technical measures. It was coded if these are mentioned in GDPR-related coverage and if so, it was investigated if it was possible to categorise these within our classes of technologies. The classes of technologies are described in detail in section 3 of deliverable D3.1.

The following tables provide an overview of the coded issues in GDPR-related coverage and the respective percentages. Table 4 shows the general topics in GDPR-related discourse in Germany and the UK, Table 5 **Error! Reference source not found.**the topics related to the consequences of the GDPR addressed in the discourse and Table 6 the assessments in the German and UK media with respect to the GDPR's long-term effects.

	Germany	**UK**
General information on the GDPR	32.26%	22.38%
Difficulties for small businesses	28.23%	13.29%
The role of Facebook & Co.	39.52%	33.57%
The Cambridge Analytica scandal	19.35%	16.08%
Data as the currency of the 21st century	11.29%	10.49%
The GDPR is not the main topic	31.45%	47.55%

Table 4: General topics in the GDPR-related discourse in Germany and the UK

	Germany	UK
Restrictions on the provision of services	20.31%	26.32%
A wave of cease and desist letters	37.50%	15.79%
Countless re-subscribe e-mails	29.69%	64.91%
Ban of social media on business smartphones	6.25%	1.75%
Europe competitiveness	31.25%	22.81%

Table 5: Topics related to the consequences of the GDPR addressed in Germany and the UK

	Germany	UK
Positive assessment of long-term effects	40.00%	15.38%
Negative assessment of long-term effects	40.00%	76.92%

Table 6: Assessments in media with respect to the GDPR's long-term effects in Germany and the UK

General issues

General information on the GDPR

Roughly one third (32.3%) of the German and one fourth (22.4%) of the UK sample contain general information on the GDPR. This was coded when the article dealt with when it would become effective, who is affected and what might happen if someone is not compliant. These articles were found most frequently shortly before and on 25 May 2018. We quickly summarise the facts that are mentioned most frequently in this regard.

The GDPR is a replacement for the 1995 Data Protection Directive, which has until now set the standards for processing data in the EU. It is frequently mentioned that the GDPR applies when a data controller or a data processor is based in the EU, but it also addresses the transfer of data outside the EU if the data collected or processed is from individuals based in the EU. Thus, many companies apply their terms globally. Moreover, it is highlighted that the regulation aims to significantly strengthen a number of rights: individuals find themselves with more power to demand companies reveal or delete the personal data they hold; regulators are able to work in concert across the EU for the first time, rather than having to launch separate actions in each jurisdiction. A fact that is mentioned in almost every article containing general information on the GDPR is that the maximum fine can now reach €20 million or 4% of the company's yearly global turnover (whichever is higher). It is stated that the GDPR affects every company, but in particular those that hold and process large amounts of consumer data: technology companies, marketers, and the data brokers who connect them. If these companies previously did not have had the tools for collating all the data they hold on an individual, it may be difficult to comply even with the basic requirements. Obviously, the largest impact will be on companies whose business models rely on acquiring and exploiting consumer data at scale. What is mentioned frequently in this regard is the meaning of consent: If companies rely on consent to process data, that consent now has to be explicit and informed - and renewed if the use changes. The world's largest companies in this regard such as Apple, Google and Facebook but also every other company had to update their sites to comply with the GDPR. Articles highlight that consumers now can request access to their personal information from data brokers or delete their information from sites altogether. However, it is also stated frequently that the regulation will have to be interpreted by court over the years, as many wordings are imprecise und ambiguous.

An issue that solely appears in the UK is 'GDPR and Brexit': What will happen to the regulation after Brexit? The regulation will shortly be part of UK law, thanks to the data protection bill that needed to be adjusted and has been working its way through parliament since September 2017, and the government has

committed to maintaining it following Brexit. In theory, a future government could change the law again - but even then, any British company wishing to do business with EU citizens would have to follow the regulation. The meaning of Brexit for data protection laws in the UK is referred to in 11.9% of the British sample.

Difficulties for small businesses

A discourse that is far more prominent in the German than in the UK coverage (28.2% vs. 13.3%) is difficulties that small businesses and other organisations like local authorities, clubs and freelancers face due to the new regulation. An investigation of the respective sources shows that while this discourse is absent in the *Guardian's* coverage, it is as prominent in the *Daily Telegraph* as it is in the German sources overall. This discourse is based on the idea that while large companies will normally have the resources to prepare for the GDPR to an appropriate extent, it will be hard for small companies that may not have the funds for extensive legal advice. At the same time, these will be hit hardest by high fines and may stop operating in fear of those.

Within this discourse, the GDPR is sometimes referred to as a "brute force"[222] that will entrench technology monopolies, as smaller apps and websites may stop operations to dodge fines. Moreover, it is stated that the GDPR will stifle smaller technology start-ups, while favouring larger players that have the funds for legal advice. Additionally, it is highlighted that new start-ups may find it hard to persuade users to consent to wide-ranging data harvesting, but if a company such as Facebook offers a take-it-or-leave-it deal, it could rapidly gain consent from millions of users. Another term that is used frequently in this regard is "bureaucratic monster"[223], putting emphasis on the expected effort to get consent, the fear of forms and fines. Especially in the German source *Süddeutsche Zeitung,* this discourse is highly visible and often expressed by features of the owners of small businesses, people who work in an honorary capacity and bloggers who work on their own. The authors share their difficulties to comply with and personal views of the GDPR.

As already mentioned, this discourse does not manifest itself as much in the British sample. Interestingly, a *Daily Telegraph* article even mentions counter advise that indeed small business were found to be more positive than negative with their views of the GDPR: According to the Daily Telegraph Business Tracker, which examined the Twitter posts of 25,000 UK companies and business people between 24 April and 27 May 2018, Britain's small business community was characterised by positivity, said to derive from an abundance of advice about the GDPR and confidence about exporting. The tracker analysed 36.000 tweets sent over the period, discovering 19% contained optimistic or cheery sentiment, while 9% were negative. 72% were neutral without any opinion attached.[224]

---

[222] BOX chief Aaron Levie in Murphy, Margi; Titcomb, James: "GDPR will force smaller sites to shut, says Box chief Levie"; The Daily Telegraph 25 May 2018.

[223] "Bürokratiemonster" in German coverage found e.g. in Strathmann, Marvin: „Datenschutz als Amazon-Bestseller; Besser als Frank Schätzing: Eine bayerische Broschüre zur neuen EU-Verordnung gehört zu den am häufigsten bestellten Büchern des Onlinehändlers", Süddeutsche Zeitung 25 May 2018; Knaut, Katharina: "Vom Aufwand für das Einverständnis; Die neue Datenschutzgrundverordnung bereitet Einrichtungen, Verbänden und Vereinen im Landkreis viel Mehrarbeit. Trotzdem werden die Regelungen als notwendig anerkannt", Süddeutsche Zeitung 29 May 2018; Zirlik, Michael: "DATENSCHUTZ; Gutes Monster", Süddeutsche Zeitung 21 June 2018.

[224] Caines, Matthew: "Full marks for SME confidence, as business owners are upbeat on GDPR advice and export prospects; Business Tracker Small and medium-sized enterprises remain positive about growth, but the TSB IT crisis did create waves", The Daily Telegraph 31 May 2018.

The role of Facebook & Co.

An issue that was already quite visible in the tag clouds (see Figure 11 and Figure 12) is the reference to Facebook within GDPR-related coverage. This reference was found in 39.5% of the German and 33.6% of the UK sample, mainly deriving from the company being found to be privacy intrusive and routinely ignoring or bending data protection law. Facebook often seems to stand for the US technology giants but other companies such as Apple, Google and Microsoft are also mentioned, although to a lesser extent.

These articles deal with the companies being privacy intrusive. Although coverage writes about Facebook launching a range of tools to officially put people in more control over their privacy, by unifying its privacy options and building an 'access your information' tool to let users find, download and delete specific data on the site, it is highlighted that the company forced every user to agree to new terms of service, and took the opportunity to nudge them into opting-in to facial recognition technology. Apple revealed a privacy dashboard of its own while noting that, unlike its competitors, it does not collect much personal data in the first place and so did not need to change much to comply. According to the authors, Google took a different tack, quietly updating its products and privacy policies without drawing attention to the changes.

According to an article in the *Daily Telegraph*[225], the Norwegian Consumer Council has issued a report warning that Internet companies continue to "deceive by design" despite claims they are compliant with new data protection law. The organisation analysed Facebook, Google and Microsoft's Windows 10 throughout April and May and found that default settings were "privacy intrusive", had misleading wording to give users an 'illusion of control' and hid privacy-friendly choices. According to the findings, the architecture of Facebook's social network, along with Google's Android operating system and various products including Gmail, made it more difficult for users to share less of their personal information, making use of so called 'dark patterns'. Dark patterns describes a technique used to mislead users through exploitative nudging, by leading someone toward making certain choices by appealing to psychological biases. Furthermore, the council found that ahead of the introduction of the GDPR visitors to Facebook's and Google's products like Gmail and YouTube were shown pop-ups threatening users with loss of functionality or deletion of the user account if the user does not choose the privacy intrusive option. For example, Facebook gives the user an impression of control over use of third party data to show adverts, while it turns out that the control is much more limited than it initially appears. According to the author, the council claimed that Google's privacy dashboard promises to let the user easily delete user data, but the dashboard turns out to be difficult to navigate, not functioning as a real tool.[226] A *BBC* article on dark patterns summarises that all three companies

- hide privacy-friendly choices;
- offer take-it-or-leave it choices;
- use privacy-intrusive defaults with a longer process for users who want privacy-friendly options
- obscure some privacy settings;
- use pop-ups compelling users to make certain choices, while key information is omitted or downplayed;
- offer no option to postpone decisions; and

[225] Murphy, Margi: "Facebook and Google breaking data laws, says consumer group", The Daily Telegraph 27 June 2018.
[226] Ibid.

- threaten users with a loss of functionality or deletion of the account if certain settings are not chosen.[227]

In this regard, within this discourse, Facebook is referred to as the ”tip of the iceberg”[228].

Other articles that reference Facebook deal with the introduction of the new ’age control’, being criticised as not offering real control, with the judgement that people who own a Facebook fanpage are in charge of data protection and also with the California Consumer Privacy Act, that is introduced in the US state and designed to provide new protections to the state’s 40 million residents in the wake of major privacy breaches including the Cambridge Analytica scandal.

The Cambridge Analytica scandal

Another very important reason for Facebook showing so prominently in our samples is the Cambridge Analytica data scandal. This was coded as a separate category, as it appeared in roughly half of the coverage referencing Facebook and thus 19.4% in the German and 16.1% in the British sample in total. In short, the personal data of tens of millions of people harvested by Facebook was shared with the political consultancy Cambridge Analytica from 2014 on and used to influence voter opinion, including the campaigns of Donald Trump and the Brexit vote. Facebook admitted that the data of 87 million users may have been improperly shared. As a consequence, on 22 May 2018 Mark Zuckerberg appeared in a hearing before a European parliament committee. This meeting is portrayed critically in our sample. Media coverage highlighted that the format allowed Zuckerberg to evade questions and give vague answers. He spent around 30 minutes giving answers to a 60-minute block of consecutive, overlapping questions that allowed the Facebook boss to pick and choose his answers. According to the respective articles, questions were ignored on shadow profiles, sharing data between WhatsApp and Facebook, the ability to opt out of political advertising, and the true scale of data abuse on the platform.

Data as the currency of the 21st century

Another discourse that we coded as part of the general issues in GDPR related coverage is what we call *‘data as the currency of the 21st century’*. It appears in 11.3% of the German and 10.5% of the UK coverage. In short, data has been described as the ”new oil” of the global economy.[229] Data brokers play a huge role in extracting value from personal information in all its forms by collecting it from hundreds of sources, including census information, surveys, public records and loyalty card programs, and by selling it to other organisations. All trade is increasingly reliant on data flows, and data brokerage is one of the fastest growing sectors of the economy.[230] While this industry seemed to be largely invisible and unregulated, the introduction of the GDPR forces data controllers to ensure data subjects understand how their information is being used.[231] In an article by *The Guardian*, security professional Bruce Schneier writes, “Surveillance is the business model of the internet. It’s not just the big companies like Facebook and

[227] “Facebook and Google use ’dark patterns’ around privacy settings, report says”, BBC News 28 June 2018.
[228] Hern, Alex: “Mark Zuckerberg appears before European parliament - as it happened; Facebook’s co-founder will be speaking to the ’conference of presidents’ made up of leaders of the eight major political groupings”, The Guardian 22 May 2018; Lynn, Matthew: “There’ll be a price to pay for privacy”, The Daily Telegraph 26 May 2018.
[229] “Facebook scandal: Who is selling your personal data?”, BBC News 11 July 2018.
[230] Crisp, James: “UK to mirror EU data rules after Brexit to secure bespoke treaty”, The Daily Telegraph 24 May 2018.
[231] “Facebook scandal: Who is selling your personal data?”, BBC News 11 July 2018.

Google watching everything we do online and selling advertising based on our behaviors; there's also a large and largely unregulated industry of data brokers that collect, correlate and then sell intimate personal data about our behaviours."[232]

An important part of that discourse is that the trade is not only done unwillingly by data brokers, but that consumers give their data in exchange for free services willingly. This can be exemplified by the use of smart home devices. While users may be comfortable trading their data for free services, or for a better quality of product, it is suggested that it may be worth thinking twice about intrusive monitoring connected to paid products - particularly when 'dumb' devices are frequently cheaper, and just as useful. A Guardian article mentions an investigation by *Which?* magazine that states that a range of connected appliances - increasingly popular features of the so-called smart home - send data to their manufacturers and third-party companies, in some cases failing to keep the information secure. The findings, according to the articles, have alarmed privacy campaigners, who warn that consumers are unknowingly building a 'terrifying' world of corporate surveillance. However, *Which?* states that data collection can also bring real benefits for those who want a more personalised service, but that consumers need to know what they are getting into when they choose to buy an Internet-connected product over a traditional 'dumb' one. With services such as Facebook and Gmail, it is obvious to get a free resource in exchange, at least partially, for access to the user's data.[233]

The GDPR is not the main topic

Especially in the very first days of our period of research and also from the middle of June on, there are many articles that only refer loosely to the GDPR, but do not deal with it directly, thus have another main topic and intention. This reflects the declining interest in the GDPR after the new regulation became effective on 25 May. Articles assigned to this category make up 31.5% of the German and 47.6% of the UK sample. There is a variety of topics covered by these articles: From sports articles that refer to the GDPR only to compare the number of scores with the number of GDPR pop-ups on the Internet over parents putting images of their children online to coverage on the financial markets.

Consequences of the GDPR

Consequences of the GDPR are portrayed in 50.6% of the German and 39.9% of the UK sample.

Restrictions on the provision of services

From the articles that contain information on the consequences of the GDPR, 20.3% in the German and 26.3% in the UK sample refer to the limitation of access of EU users to certain international services. Prominent examples include the newspapers owned by Tronc Inc. (LA Times, Chicago Tribune, New York Daily News, the Baltimore Sun, Orlando Sentinel and the San Diego Union-Tribune), that redirect users to a page saying that the respective website is currently unavailable in EU countries. Other services have announced that they are permanently deleting the accounts of EU-based users in order to comply with the internationally applicable regulation. At the end of June, the Pinterest-owned reading app Instapaper

[232] Schneier, Bruce: "Data protection laws are shining a needed light on a secretive industry; Regardless of where we live, we all benefit from data protection laws - companies must us show how they profit off our information", The Guardian 1 June 2018.
[233] Hern, Alex: "UK homes vulnerable to 'staggering' level of corporate surveillance; Smart home appliances send data to manufacturers and third parties, Which? warns", The Guardian 1 June 2018.

was down for maintenance for a month; and USA Today took steps to offering those in the EU a slimmed-down, ad-free experience, hoping that will leave their title compliant with the law.

Figure 13: Los Angeles Times is still blocked to EU readers (screenshot)

While most websites, according to the articles in the sample, stressed that the cut-offs were temporarily while implementing the required technical compliance solutions, most of them are still not available to EU readers.[234]

A wave of cease-and-desist letters

The discourse on cease-and-desist letters differs in the two countries: While it is very prominent in the German reporting that deals with the consequences of the GDPR (37.5%), it is much less frequently found in the respective British coverage (15.8%) and also has another focus. This mainly derives from Germany having a long "tradition" of certain lawyers making money from dissuasions that refer to minor formal errors of internet companies.

In the German coverage, efforts by political parties to prepare for a wave of cease-and-desist letters were described. Such letters were thought to come due to judicial insecurity caused by the GDPR. Some lawyers expected other advocates to try to make profit from cease-and-desist letters. The German federal government already planned a law against these, referring to the past when complicated new regulations concerning cyber law have provoked such waves of cease-and-desist letters, mainly to protect SMEs, freelancers and associations from fines. However, after a certain time it seemed to become clear that no such wave is coming, but that the cease-and-desist letters that have led to the political actions against

[234] Last checked on 7 September 2018

those probably have been sent by very few and always the same lawyers. An article in the *Frankfurter Allgemeine Zeitung* from the beginning of July refers to it as a 'phantom'.[235]

In the UK, the discourse is less prominent, and there are no political efforts mentioned as described for Germany.

Countless re-subscribe e-mails

In the weeks leading to the enforcement of the GDPR, consumers were inundated with e-mails asking them to "opt in" to continue receiving material. This issue is covered quite extensively in both the German and the British media, as it is referred to in 29.7% of the German and even 64.9% of the British articles that deal with consequences of the GDPR, constituting the by far most frequently mentioned consequence of the GDPR in the British sample.

These e-mails are based on the law that says that businesses must secure consent before e-mailing people with marketing materials. In most cases, they will already have secured this consent by making the user tick a box or press an 'accept' button at some point. According to the articles, some companies are struggling because they have not actually kept a record of whether or not they have done this.[236] However, most people seem to ignore those e-mails and there are articles that even refer to a "once in a lifetime opportunity" to get rid of all the unwanted newsletters.[237] Additionally, consumers seem to even more ignore the e-mails as the tone of them gets increasingly desperate, feeling they were being "emotionally blackmailed" into replying.[238] Some small businesses are reporting that 'reconfirmation' rates are averaging just 10%, meaning they are losing 90% of their marketing e-mail lists.

A part that is covered in the British and almost absent in the German sample is the fact that these e-mails might even be illegal: Some legal experts argued that many of these e-mails are unnecessary, and may even be illegal under existing data protection laws. If a consumer has previously expressly consented to receive e-mails from a company, that consent would remain valid under the new legal framework. According to a *Guardian* article, there are also five other justifications for companies to send e-mails to individuals whose personal data they possess - contract, legal obligation, vital interests, public interest and legitimate interests.[239] Thus, if a company does not hold legal consent, it is not allowed to send e-mails to individuals at all.

Companies have generally sorted in one of two camps, depending on what legal advice they've taken. On the one hand are those who argue they have a "legitimate interest" in processing personal data, and just feel the need to notify their customers of the forthcoming changes to their terms and conditions; on the other hand are those who believe they need explicit consent from their customers to keep in touch. Either

[235] Wieduwilt, Hendrik: "Datenschutz-Abmahnwelle bislang nur ein Phantom; Bisher kaum Fälle bekannt / Justizministerium plant keine spezifische Abmahnbremse", Frankfurter Allgemeine Zeitung 3 July 2018.
[236] Samuel, Juliet: "All these data emails are totally unnecessary", The Daily Telegraph 26 May 2018.
[237] Collinson, Patrick; Jones, Rupert: "NHS warns patients they could lose text alerts as GDPR deluge continues; Health service joins UK firms in rushing to comply with new data protection rules", The Guardian 19 May 2018.
[238] Morley, Katie et al.: "Firms accused of 'emotional blackmail' over new data law; Consumers 'bribed' to stay on email lists", The Daily Telegraph 25 May 2018.
[239] Belam, Martin: "Businesses resort to desperate emailing as GDPR deadline looms; As regulations come into force on Friday, inboxes fill with messages hoping to persuade customers to stay subscribed", The Guardian 24 May 2018.

way, the worst-case scenario is usually that ignoring an e-mail will mean that an individual receives fewer in the future.

Another aspect that is more prominently found in the British articles concerns the potential of phishing attacks that comes with these e-mails: Extortionists are already using spoof e-mail templates to lure people into inadvertently sharing their personal or financial information, cyber security experts have warned. In one instance, which was taken up by the media, a fake GDPR e-mail claiming to be from Airbnb attempted to trick users into clicking on a link that would have asked them to input credit card information.[240]

In both countries, however, the sending of GDPR e-mails is framed as an opportunity that should not be missed. Or, as the *Guardian* puts it, "The GDPR is a rebalancing of power between us, the people who have to hand over data to do transactions on the internet, and the organisations that intend to blitz us into submission with emails. (...) So, sure, those emails are a bit of a pain, and a laugh. But they're also part of a long-overdue recognition that companies have been too lazy about their security with our data. It's exactly the data detox that we all need."[241]

Ban of social media on business smartphones

An issue that seemed to be prominent in the media reporting during our exploratory phase in the beginning of June was that automotive giant Continental had banned workers from using WhatsApp and Snapchat on their smartphones over concerns the apps might be leaking confidential information to third parties. The ban, which is effective for 240,000 employees in 61 countries, follows the introduction of the GDPR, as it gives organisations that collect and store personal information more liability in the event of a breach.[242] The company feared that the apps have deficiencies when it comes to data protection, as they access a user's personal and potentially confidential data such as contacts, and thus the information of third parties who are not involved. The company criticised that the apps handed over the responsibility to the users, who theoretically would need to get consent for data sharing from every single contact. The company however is cited that the ban could be released immediately if the providers would set up their services to be GDPR compliant by default.[243]

However, the issue actually is mentioned in only 6.3% of the German articles that deal with consequences of the GDPR and even only 1.8% of the respective British ones and thus does not build a prominent part of the GDPR-related discourse in both countries.

An interesting detail, however, is that in Germany the initial reports from the beginning of June are followed by coverage on other companies that plan to ban certain social media apps as well (namely BASF, Beiersdorf and ZF Friedrichshafen). German software giant SAP is said to give its employees permission to still use these services, but to be working on an internal solution. Another mentioned solution is the Swiss service Threema, which, on the one, hand is a paid app an thus has less interest to share user data

[240] Murphy, Margi et al.: "Cyber criminals send spoof mail to exploit privacy policy updates", The Guardian 15 May 2018.
[241] Arthur, Charles: "Liberation day! Don't email me. I sure won't be emailing you; Today, happily, the EU's General Data Protection Regulation comes into force. It's the data detox we've been waiting for", The Guardian 25 May 2018.
[242] Murphy, Margi: "End of the road for WhatsApp as Continental reveals ban", The Daily Telegraph 6 June 2018.
[243] Jauernig, Henning: "Continental verbietet WhatsApp und Snapchat auf Diensthandys", Spiegel Online 5 June 2018.

inappropriately, and, on the other hand, offers a special messenger meant for businesses, which is said to be 100% GDPR compliant.[244]

European competitiveness

An issue that is of special interest in our research is the question if the media would portray possible consequences of the GDPR with regard to the future role of Europe in the international context: Are there any assessments of the long-term effects for the EU and EU based companies? And if so, are the new data protection regulations seen more positively or negatively? Might privacy even be a competitive advantage?

Our research shows that indeed there was such a discourse, as it was found in 31.3% of the German articles that deal with the consequences of the GDPR and 22.8% of the respective British ones. What is remarkable, though, is that within these articles, in the German media the assessment was found to be quite balanced: Both positive and negative consequences are mentioned in 40% of the articles, 20% of the articles are neutral (i.e. no opinion attached). In the respective British articles, only 15.4% mention positive consequences for the EU while 76.9% see the more negative side with 7.7% being neutral.

The German articles that feature positive consequences address a wide range of different aspects. Though certain weaknesses are acknowledged, the GDPR is referred to as "the best protective shield for citizen's privacy"[245], resulting in the big technology companies having changed their data protection settings. Other countries would try to import European data protection regulations, which is seen as a demonstration of power of the EU in the digital age. If other states do not want to lose EU customers and threaten their relationship with the EU, they have to change their habits of data processing.[246] Charles-Édouard Bouée, Chief Executive Officer of consultancy Roland Berger, states in a *Frankfurter Allgemeine Zeitung* article that with the GDPR, the EU has set the standard for data protection and that its leading the way will enable other technological innovations.[247] In a *Süddeutsche Zeitung* article, German politician and driving force behind the GDPR Jan Philip Albrecht, is cited to not having expected that the EU would set the global standard for data protection one day, and that this standard is aimed to be copied by American and Chinese companies.[248] In another article by the same author, it is stated that the EU could have any right to be proud, as the new regulation does not only protect the data of every EU citizen, but is also widely seen as a foundation pillar of a developing global standard of data protection.[249] Even in the US, where data protection has previously been seen as barrier to innovation, the regulation is now regarded as an inspiration, says the European Commissioner for Justice, Consumers and Gender Equality Věra Jourová.[250] "The *New World* must learn from the *Old World"*, American politician and former chairman of the Federal Communications Commission Tom Wheeler is cited in the same article, expressing hope that the

[244] Astheimer, Sven et al.: "Arbeitgeber wollen Whatsapp und Co. nicht; Nicht nur Continental verbannt die Nachrichtendienste aus Daten-schutzgründen von den Diensthandys", Frankfurter Allgemeine Zeitung 6 June 2018.
[245] "der beste Schutzschirm für die Privatsphäre von Bürgern, den es je gab" (Brühl, Jannis: "INTERNETKONZERNE; Leichtes Spiel für Zuckerberg", Süddeutsche Zeitung 23 May 2018).
[246] Ibid.
[247] Schubert, Christian: "'Europa muss seine Bürger schützen'", Frankfurter Allgemeine Zeitung 15 June 2018.
[248] Kirchner, Thomas: "NAHAUFNAHME; Vater der DSGVO; Der Grünen-Politiker Jan Albrecht geht von Brüssel nach Kiel", Süddeutsche Zeitung 2 July 2018.
[249] Kirchner, Thomas: "Plötzlich wird es ernst. Nach diversen Skandalen sieht sich die EU in ihren strengen Regeln bestätigt. Sogar US-Experten gelten sie als Vorbild", Süddeutsche Zeitung 24 May 2018.
[250] Ibid.

regulation would have a positive impact on global data protection habits.[251] The US is always closely linked to the big technology companies, that previously have set the rules, but that are now forced to rethink their services.[252]

While this discourse is also found in the UK coverage, it is far less prominent and less detailed. "While certainly not perfect, this is the biggest ever attempt by governments to get to grips with the huge new powers that digital technology hands to big business. (...) GDPR (General Data Protection Regulation) is one part of the answer, passed in the face of fierce opposition from big tech" Nick Dearden of Global Justice Now writes in a *Guardian* article.[253] Bruce Schneier also sees the global benefits for data privacy and security, as personal data can only be collected and saved for specific purposes and only with the explicit consent of the user, eventually resulting in forcing the big technology companies to show how they profit from European users' personal data.[254]

In a *Süddeutsche Zeitung* article, the results of a survey by the German trade association Bitkom are cited. This survey, according to the article, shows that German companies' expectations are quite balanced regarding positive or negative consequences for their businesses. According to the survey, seven out of ten companies expect advantages concerning uniform competitive conditions in the EU, with 43% expecting advantages for their very own business. However, also 50% of the companies fear that processes might become more complicated and expect additional work and expenses. 38% fear that the GDPR will pose a barrier to innovation and digitization in Europe.[255]

As initially mentioned, portrayal of the potential negative consequences for Europe does take place in the German reporting on the GDPR to a similar extent. However, when looking at the sources in detail, it becomes apparent that these are only found in the conservative *Frankfurter Allgemeine Zeitung*. Correspondingly, the vast majority of the articles that portray potential negative consequences in the UK sample are published by the also conservative *Daily Telegraph*. Within those articles, the GDPR is generally portrayed as going to hurt the European economy in different ways.

Arguments presented in both countries' coverage include expected additional expenses for the whole of Europe. According to the respective articles, however, it is not foreseeable how much extra costs the "data bureaucracy" brings to Europe as a business location. Examples for things that bring extra costs are inefficient services, the need for extensive documentary and the consultation of lawyers.[256] According to an article by *The Daily Telegraph*, an American estimate for costs is around $7.8 billion for the Fortune 500 companies alone, which, according to the author, probably needed to be at least doubled for a global

[251] Ibid.
[252] Kirchner, Thomas: "URHEBERRECHT; Diebe im Netz", Süddeutsche Zeitung 6 July 2018; Karliczek, Anja: "STANDPUNKT; Für eine neue Datenpolitik", Frankfurter Allgemeine Zeitung 26 May 2018.
[253] Dearden, Nick: "Big tech companies are trying to rewrite the rules to get your data; Regulations such as the EU's GDPR may not frighten the Silicon Valley giants, but they're a step in the right direction", The Guardian 22 May 2018.
[254] Schneier, Bruce: "Data protection laws are shining a needed light on a secretive industry; Regardless of where we live, we all benefit from data protection laws - companies must us show how they profit off our information", The Guardian 1 June 2018.
[255] DPA: "Viele Firmen nicht bereit für neuen Datenschutz", Süddeutsche Zeitung 18 May 2018.
[256] Wieduwilt, Hendrik: "Preis des Datenschutzes", Frankfurter Allgemeine Zeitung 5 June 2018.

figure. “Wages will be lower, dividends will be reduced, prices will be raised, and factories won’t get built as money is diverted to data compliance instead.”[257]

Another argument found in both countries’ coverage is closely linked to the category ‘difficulties for small businesses’: In contrast to big corporations that have the resources to spend a lot of time and money on the changes needed, small companies will be hit hard, which will eventually contribute to a further imbalance of power.[258] The GDPR is referred to as being a brake (“Bremsklotz”) for the European economy[259], as an exaggerated barrier[260], that might lead to a significant slow-down of digitization in Europe.[261] Europeans are described as throwing sand into the wheels of European Internet services,[262] which the EU cannot afford. New technologies such as driverless cars and AI systems will need to collect massive amounts of data and share it online to function properly, and with such regulations, it is seen as unlikely that such technologies will be developed in Europe in the future.[263] China, for instance, is mentioned to benefit from having free access to large amounts of data, essential for training algorithms, in contrast to Europe where the introduction of the GDPR is referred to as a potential drawback in Europe.[264]

Some articles refer to the ePrivacy law, referring to the GDPR as ”the tip of the iceberg”[265] that will be followed by even more restrictive legislation that will hamper the European creative industry even more.[266] A letter to the European Council signed by 57 technology and Internet groups warns of a “considerable negative impact” that will “extend to all sectors of the EU digital economy”, saying that the proposed regulations are overly strict and will prevent legitimate data processing, even when there is little risk of a person’s privacy being breached. For example, communications between machines would be covered, potentially hampering new technology such as driverless cars. “The ePrivacy proposal has departed from the laudable objective of protecting the confidentiality of communications and goes on instead to greatly limit the processing of a broad array of both personal and non-personal data,” the letter to the EU Council says. “The considerable negative impact of an inflexible ePrivacy [law] will extend to all sectors of the EU digital economy - from digital media to connected cars, medical technology and smart

[257] Lynn, Matthew: “There'll be a price to pay for privacy”, The Daily Telegraph 26 May 2018.
[258] Ibid.; Thiel, Thomas: “Worauf sich Sisyphus jetzt konzentrieren muss; Martin Schallbruch führt vor, wie sich der überforderte Staat erfolgreich in der Digitalisierung behaupten kann”, Frankfurter Allgemeine Zeitung 29 June 2018
[259] Ibid.
[260] Klimm, Leo: “Auf Krawall gebürstet; 'Wir mögen das ewige Palaver nicht': Beim OECD-Treffen in Paris prallen die Vorstellungen der Amerikaner und die der Europäer über die künftige Ordnung der Weltwirtschaft aufeinander”, Süddeutsche Zeitung 1 June 2018.
[261] Lindner, Roland: “NETZWIRTSCHAFT; Acer fürchtet die Digitalisierungsbremse. Der Computerhersteller fürchtet die Schattenseiten der Datenschutzgrundverordnung und drängt stärker in die Schulen”, Frankfurter Allgemeine Zeitung 28 May 2018.
[262] Wieduwilt, Hendrik: “Preis des Datenschutzes”, Frankfurter Allgemeine Zeitung 5 June 2018.
[263] Lynn, Matthew: “There'll be a price to pay for privacy", The Daily Telegraph 26 May 2018.
[264] "Alibaba reveals new driverless delivery bot”, BBC News 1 June 2018.
[265] Lynn, Matthew: “There'll be a price to pay for privacy”, The Daily Telegraph 26 May 2018.
[266] Rabe, Thomas: “Für eine wettbewerbsfähige Regulierung; Europas Kreativindustrie darf im Wettbewerb mit Google und Facebook nicht benachteiligt werden” Frankfurter Allgemeine Zeitung 4 July 2018.

manufacturing - which will be exposed to additional burden at best or, at worst, unable to continue offering and innovating their products and services using data."[267]

Technical solutions to act in a GDPR-compliant way

Lastly, we paid some attention to technical measures that companies can implement to provide their services in a GDPR compliant way. It was coded if these are mentioned in GDPR related coverage and if so, it was investigated if it was possible to categorise these within e-SIDES' classes of technologies. In short, these are the following[268]: anonymisation, sanitisation, encryption, deletion, multi-party computation, access control, policy enforcement, accountability, data provenance, transparency, access and portability, and user control.

Anonymisation is performed by encrypting or removing personally identifiable information from datasets. Traditional anonymisation techniques fail in the context of big data applications because there are too many data points for a single individual. **Sanitisation** is done by encrypting or removing sensitive information from datasets. **Deletion** is closely related to sanitisation but focuses on the permanent erasure of data from a physical medium. Anonymisation is a type of sanitisation. **Encryption** is the encoding of information so that only authorised parties can access it. In the context of big data applications, it is necessary to go beyond the "encrypt all or nothing" model. **Multi-party computation** is a field of cryptography that relies on the distribution of data and processing tasks over multiple parties. Although it was proven to be theoretically plausible, there are no practical solutions yet.

Access control describes the selective restriction of access to places or resources. Big data applications typically require fine-grained access control and traditional approaches increasingly become unmanageable. **Policy enforcement** focuses on the enforcement of rules for the use and handling of resources. Automated policy enforcement mechanisms are considered particularly important in the big data era. **Accountability** requires the evaluation of compliance with policies and the provision of evidence. A cornerstone of accountability in the context of big data applications is the provision of automated and scalable control and auditing processes. **Data provenance** relies on being able to attest the origin and authenticity of information. The aim is to provide a record of the processing history of pieces of data.

Transparency refers to the provision of intelligible and easily accessible information regarding the approach and algorithms for data collection and processing. It may be achieved by providing purely textual information, through multichannel and layered approaches, or standardized icons and pictograms. **Access and portability** facilitates the use and handling of data in different contexts. Having access to data means that data subjects can look through and check the data stored. **Portability** gives data subjects the possibility to change service providers without losing their data. **User control** refers to the specification and enforcement of rules for data use and handling. Means that allows reaching user control include consent mechanisms, privacy preferences, sticky policies and personal data stores.

As the project's aim is to complement the research on privacy-preserving big data technologies and data-driven innovation, we are particularly interested in methods that enable organisations to act in a GDPR-

[267] Titcomb, James: "Tech firms slam 'careless' EU privacy rules; Companies react to plans for EU crackdown on reading private messages between individuals", The Daily Telegraph 11 June 2018.

[268] Bachlechner, Daniel, Michael Friedewald, Jana Weitkamp, and Nicholas Martin, "Overview of existing technologies", e-SIDES Deliverable 3.1 e-Sides Project, 2018. https://e-sides.eu/assets/media/e-SIDES%20D3.1%20V1.0.pdf

compliant way. In this regard, the e-SIDES project already defined classes of technologies that may have the potential to address some of the ethical, legal, societal and economic issues that are raised by big data applications. We have examined if the technologies and technology-related needs mentioned in the GDPR-related coverage can be mapped to the e-SIDES classes of technologies.[269]

Due to the particular relevance of privacy issues in the context of new ICTs, it has been tried to address the issues, among others, by means of technological measures. One way is to create new technologies that address issues of the previous technologies. A typical example of such a technological solution are anonymisers. These are technological tools to anonymise personal data and aim to mitigate privacy issues.

Although such technological solutions may be helpful in addressing the issues, the e-SIDES project does not focus on these types of technological solutions but rather on a new approach, in which new technologies are designed in ways that they actually are already privacy-preserving when they are ready for their first-time use. The idea is to distil design requirements from ethical, legal societal and economic perspectives and to build new technologies that take these requirements into account. In other words, privacy requirements are taken into account throughout the entire engineering process. This is what is called Privacy by Design[270] and became mandatory through Art. 25 GDPR.

Accordingly, we researched if articles dealing with the GDPR might explicitly call out such technologies. Indeed, we found that 26.61% of the German and 29.37% of the British sample contained references to the need for such technologies. **Error! Reference source not found.** provides an overview which technologies are addressed most often within these articles. As in both countries there were no references to anonymisation, multi-party computation and data provenance, these classes are not included in Table 7.

[269] Bachlechner, Daniel, Michael Friedewald, Jana Weitkamp, Melek Akca Prill, Karolina La Fors, and Alan M. Sears, "Assessment of existing technologies", e-SIDES Deliverable 3.2 e-Sides Project, 2018. https://e-sides.eu/resources/deliverable-d32-assessment-of-existing-technologies

[270] Ann Cavoukian and Jeff Jonas, "Privacy by Design in the Age of Big Data", (Information and Privacy Commissioner, Ontario, Canada, 2012), https://jeffjonas.typepad.com/Privacy-by-Design-in-the-Era-of-Big-Data.pdf (accessed December 14, 2017)

Technology	Germany	UK
Encryption	18.18%	16.67%
Deletion	42.42%	26.19%
Access control, policy enforcement	33.33%	11.90%
Accountability, audit mechanisms	9.09%	4.76%
Transparency	57.58%	35.71%
Access and data portability	57.58%	33.33%
User control	21.21%	88.10%

Table 7: References to technology classes in Germany (n=33) and the UK (n=42)

While there are frequent references in both Germany and the UK to deletion, transparency, access and data portability and user control, references to encryption are less frequent. The least references in both countries are found to accountability and audit mechanisms. In the UK coverage, user control stands out with 88.1%, whereas in Germany, references to both transparency and access and data portability are found most often with percentages of 57.58%.

User control is so prominently mentioned in our sample because it is closely related to consent. Articles stress that the GDPR requires organisations to obtain explicit consent before processing personal data. For example, consumers must agree to have their data collected, shared and used for targeted advertising. The references to user control mechanisms are also strongly connected to the discourse "Countless re-subscribe e-mails" because in these articles it is often stressed that positive opt-in and informed consent are needed.

Providing users with access to their data is an important privacy condition as well as an obligation of data controllers. Data portability is an additional and important aspect for users as it gives them the possibility to change service providers without losing data. Accordingly, access and data portability are mostly mentioned in articles on companies introducing tools to access and download the data that they store about customers. Articles also mention that customers can now easily request a copy of any personal information held.

Transparency in our sample mostly refers to as the need for information to be easy to find and clearly laid out. Proper information and transparency is a key issue in any data processing, so as to allow individuals to understand how their data are being processed and to make informed choices. Accordingly, articles highlight that customers need to be able to see exactly how their information is being used. As an example, in a *Guardian* article, Facebook is cited that "[p]eople are able to manage their ad preferences tool, which clearly explains how advertising works on Facebook". People can tell if they want to see ads based on specific interests or not. When interests are removed, Facebook claims to show people the list of removed interests so that they have a record they can access, but these interests are no longer used for ads.[271]

While a key concept in data protection law, references to accountability and audit mechanisms are not frequent in our sample. Accountability refers to the provision of automated and scalable control and auditing processes that can evaluate the level of compliance with privacy policies against predefined

[271] Hern, Alex: "Facebook lets advertisers target users based on sensitive interests; Social network categorises users based on inferred interests such as Islam or homosexuality", The Guardian 16 May 2018.

machine-readable rules. However, in our sample, references to accountability are only made in the sense that organisations are said to be accountable for their customers' data and must prove they have consent.

References to access control and policy enforcement are quite frequent in our sample. Access control ensures that only authorised entities can gain access to data. In the analysed articles, the need for access control mechanisms is mostly hidden in sentences such as "data must be kept safe and secure", so that it cannot be stolen or get lost.

Deletion concentrates on the removal of sensitive information from a dataset. An efficient approach to secure data deletion may be to securely delete the key needed to encrypt properly encrypted data. In our sample, references to deletion are mostly related to companies providing tools to let users find, download and delete specific data on their sites. What is also mentioned in this regard is the "right to be forgotten" and that a customer can ask a company to delete all the data it holds about him or her.

Encryption transforms data in a way that only authorised parties can read it. Articles in our sample do not often mention encryption explicitly but hide it in the phrases such as "data must be kept safe and secure" and "high level of data protection" is needed. However, a *Guardian* article also mentions that the "new GDPR (General Data Protection Regulation) is encouraging companies to adopt secure, encrypted email services".[272]

[272] Schofield, Jack: "Can my employer read emails in my Gmail account?", The Guardian 28 June 2018.