TURUN YLIOPISTO
Oikeustieteellinen tiedekunta

The collection of personal data from parties other than the data subject under the GDPR

Immateriaalioikeudet ja informaation muu sääntely
Pro Gradu -tutkielma

Laatija:
Sami Aho

28.4.2023

TURUN YLIOPISTO
Oikeustieteellinen tiedekunta

Tiivistelmäsivu

Tutkielma

Oppiaine: Immateriaalioikeudet ja informaation muu sääntely
Tekijä: Sami Aho
Otsikko: The collection of personal data from parties other than the data subject under the GDPR
Ohjaaja: Tuomas Mylly
Sivumäärä: 62 sivua
Päivämäärä: 28.4.2023

Yleinen tietosuoja-asetus on vuodesta 2018 säädellyt henkilötietojen käsittelyä EU:n alueella vahvistaen Euroopan unionin perusoikeuskirjassakin turvattujen oikeuksien toteutumista yksityisten henkilöiden oikeudesta yksityiselämään sekä henkilötietojen suojaan. Henkilötietojen käsittelyyn osaa ottavat tahot ovat joutuneet uudelleenarvioimaan omaa rooliaan, vastuutaan sekä velvollisuuksiaan noudattaakseen asetuksen velvollisuuksia. Viime aikojen kiihtynyt kehitys tekoälyn saralla on herättänyt keskustelua yleisen tietosuoja-asetuksen yhteensopivuudesta uuden teknologian tarpeisiin ja erityispiirteisiin.

Tässä tutkielmassa tutkitaan eri tahojen, niin rekisterinpitäjän, yhteisrekisterinpitäjän kuin henkilötietojen käsittelijänkin yleisen tietosuoja-asetuksen asettamia velvollisuuksia ja vastuita sekä käsittelyn lainmukaisuuden rajanvetoa tilanteissa, joissa henkilötietoja kerätään muilta tahoilta kuin rekisteröidyltä itseltään. Tarkastelu kohdentuu erityisesti informointivelvollisuuteen sekä henkilötietojen käsittelyn lainmukaisuuteen tuoden samalla esiin yleisen tietosuoja-asetuksen henkilötietojen käsittelyn yleiset periaatteet, kuten tietojen minimoinnin periaatteen sekä tarkoituksenmukaisuusperiaatteen. Kriittisen tarkastelun alle joutuu myös itse henkilötiedon määritelmä.

Tutkielma on lainopillinen katsaus yleisen tietosuoja-asetuksen soveltamisesta, pyrkien käytännönläheiseen tulokulmaan tuoden esiin tosielämässä kohdattuja haasteita oikeuskäytäntöä hyödyntäen. Viimeisessä kappaleessa on hyödynnetty rajoitetussa mittakaavassa myös *de lege ferenda* -metodia.

Yleisen tietosuoja-asetuksen asettamat velvoitteet ja vaatimukset eivät aina vastaa tosielämän käytäntöjä. Monisyinen lainsäädäntökokonaisuus on saattanut unohtaa sille osoitetun ytimensä, yksityishenkilöiden arkaluontoisten tietojen suojaamisen, oikeushyvän, jota sen olisi suojeltava. Henkilötietojen kerääminen muilta osapuolilta kuin rekisteröidyltä itseltään saattaa olla monelle toimijalle elinehto, jolloin yleisen tietosuoja-asetuksen noudattaminen pilkun tarkkuudella on saattanut jäädä toissijaiseksi tavoitteeksi.

Avainsanat: Yleinen tietosuoja-asetus, lainmukaisuus, henkilötietojen kerääminen kolmansilta osapuolilta, tekoäly, henkilötietojen määritelmä, tietojen minimoinnin periaate, käyttötarkoitussidonnaisuusperiaate, suostumus, oikeutetut edut

TURUN YLIOPISTO
Oikeustieteellinen tiedekunta

Abstract

Master's thesis

Subject: Immateriaalioikeudet ja informaation muu sääntely
Author: Sami Aho
Title: The collection of personal data from parties other than the data subject under the GDPR
Supervisor: Tuomas Mylly
Number of pages: 62 pages
Date: 28 April 2023

Since 2018, the General Data Protection Regulation has regulated the processing of personal data in the EU, reinforcing the individuals' rights to privacy and the protection of personal data, as enshrined in the Charter of Fundamental Rights of the European Union. Those involved in the processing of personal data have had to reassess their roles, responsibilities, and obligations in order to comply with the obligations of the regulation. Recent accelerated developments in the field of artificial intelligence have prompted a debate on the compatibility of the General Data Protection Regulation with the needs and specificities of new technologies.

This thesis will examine the obligations and responsibilities of the various actors, whether they are controllers, joint controllers or processors of personal data, as well as the limits of lawfulness of data collection under the General Data Protection Regulation in situations where personal data are collected from parties other than the data subject. In particular, the review will focus on the obligation to inform and the lawfulness of the processing of personal data, while highlighting the general principles of the GDPR in relation to the processing of personal data, such as the principle of data minimization and the principle of purpose limitation. The definition of personal data is also subject to critical scrutiny.

The thesis is a jurisprudential review of the application of the General Data Protection Regulation, aiming at a pragmatic approach to the outcome, highlighting real life challenges by making use of case law. The last chapter also makes limited use of the *de lege ferenda* method.

The obligations and requirements imposed by the General Data Protection Regulation do not always correspond to real-life practices. The complex body of legislation may have lost sight of its core, the protection of individuals' sensitive data, a legal interest that it should protect. The collection of personal data from parties other than the data subjects themselves may be the lifeblood of many operators, which may have made compliance with the General Data Protection Regulation (GDPR) a secondary objective.

Keywords: GDPR, lawfulness of processing, data collection from third parties, artificial intelligence, definition of personal data, principle of data minimization, principle of purpose limitation, consent, legitimate interests

Table of contents

References ..... **VII**

Abbreviations ..... **XIV**

1 Introduction **1**

1.1 Background **1**

1.1.1 Relevance and importance of the topic 1

1.1.2 Fragmented regulation regarding the data collection 2

1.1.3 User data collection as a phenomenon 3

1.2 Research question, methods and materials **5**

1.2.1 Research questions and focus of the thesis 5

1.2.2 Methodology and materials 7

1.2.3 Limitations of the thesis 10

2 Data collection and the GDPR **11**

2.1 Scope of application of the GDPR **11**

2.2 Relevant definitions under the GDPR **11**

2.2.1 Personal data 11

2.2.2 Sensitive personal data 12

2.2.3 Data controller, data processor, and joint controllers 13

2.2.4 Data subject, third parties, recipients of data 15

2.3 Lawfulness of data processing – principles of the GDPR **16**

2.3.1 Principle of data minimization 16

2.3.2 Purpose limitation principle 16

3 Information obligations for data collector and data recipient **18**

3.1 Set-up **18**

3.2 Information obligations of the data collector under the GDPR when collecting personal data from the data subject **19**

3.2.1 Real-life data collection examples 19

3.2.2 Data collector as an independent controller 20

3.2.3 Data collector as a processor 25

3.2.4 Joint controllers 25

3.3 Information obligations of the data recipient when collecting data from other sources **27**

4 Relevance of the legal basis for data collection 29

4.1 Consent 29

4.1.1 General definitions and requirements 29
4.1.2 Obtaining and revoking consent 30
4.1.3 Detriment 32
4.1.4 Function creep and consent fatigue 33

4.2 Performance of a contract 34

4.2.1 General conditions 34
4.2.2 Limitations to the use of performance of a contract as a lawful basis 35

4.3 Legitimate interests 36

4.3.1 General conditions 36
4.3.2 Balancing test 38
4.3.3 Reasonable expectations 39
4.3.4 Legitimate interests of third parties 39
4.3.5 Additional safeguards 41
4.3.6 Data collection to train artificial intelligence 41

5 The German Facebook Case 46

5.1 Introduction, the legal issues at stake 46

5.2 Legal basis for data collection 47

5.2.1 Questions referred to the CJEU 47
5.2.2 Performance of a contract 49
5.2.3 Legitimate interests 51
5.2.3.1 Balancing test regarding the legitimate interests 51
5.2.4 Effectiveness of a consent 55

6 Conclusions 57

6.1 Outer limits of data collection through sources other than data subjects 57

6.1.1 Legal bases for data collection and the principles data processing 57
6.1.2 Are there any lessons learned from the German Facebook case? 58

6.2 Looking ahead: is the GDPR future-proof? 58

6.2.1 Big Data, AI, and the GDPR 58
6.2.2 The definition of personal data 59
6.2.3 Protecting individuals by hindering markets? Principles of data minimization and purpose limitation 61

References

Literature

Chander, Anupam, ‘Is Data Localization a Solution for Schrems II?’ (July 27, 2020). Journal of International Economic Law.

Cohen, Julie E., ‘Everything Old is New Again – Or Is It?’ in ‘Between Truth and Power: The Legal Constructions of Informational Capitalism’, Oxford Academy 2019.

Corrales Compagnucci, Marcelo – Mateo, Aboy – Minssen, Timo, ‘Cross-Border Transfers of Personal Data After Schrems II: Supplementary Measures and New Standard Contractual Clauses (SCCs)’, 30 December 2021.

Ebers, Martin, 'Regulating AI and Robotics: Ethical and Legal Challenges' in Martin Ebers/Susana Navas Navarro (eds.), Algorithms and Law, Cambridge, Cambridge University Press, 2019.

Edwards, Lilian, ‘Privacy, security and data protection in smart cities: a critical EU law perspective’ (2016) 2(1) European Data Protection Law Review 28.

Feiler, Lukas – Forgó, Nikolaus – Weigl, Michaela, ’The EU General Data Protection Regulation (GDPR): A Commentary. Globe Law and Business 2018.

Finck, Michele – Asia J. Biega, ‘Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems’. Technology and Regulation, 7 December 2021, 44-61.

George, Damian – Reutimann, Kento – Tamó-Larrieux, Aurelia. ‘GDPR Bypass by Design? Transient Processing of Data under the GDPR’. *International Data Privacy Law* 9, no. 4 (2019): 14.

Hacker, Philipp, ‘A Legal Framework for AI Training Data’ (2021) 13(2) Law, Innovation and Technology 257, p. 291.

Hintze, Mike ‘Viewing the GDPR Through a De-identification Lens: A Tool for Compliance, Clarification, and Consistency’ (2018) 8 International Data Privacy Law 86.

Hendrickx – Rocher – de Montjoye, ‘Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models’ (2019) 10 Nature Communications 3069.

Kamara, Irene – De Hert, Paul, Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach. In Seligner,

E., Polonetsky, J., and Tene, O., editors, The Cambridge Handbook of Consumer Privacy. Cambridge University Press 2019.

Kerber, Wolfgang – Zolna, Karsten K.. 'The German Facebook Case: The Law and Economics of the Relationship between Competition and Data Protection Law'. *European Journal of Law and Economics* 54, no. 2 (October 2022): 217–50.

Koops, Bert-Jaap, 'The Concept of Function Creep.' Law, innovation and technology 13.1 (2021): 29–56.

Korpisaari, Päivi – Pitkänen, Olli – Warma-Lehtinen, Eija: Uusi tietosuojalainsäädäntö. 2018 Alma Talent Oy.

Kramcsák, Pablo Trigo, 'Can legitimate interest be an appropriate lawful basis for processing Artificial Intelligence training datasets?, Computer Law & Security Review 48 (2023).

Langhanke, Carmen – Schmidt-Kessel Martin, "Consumer Data as Consideration" (2015) 4 Journal of European Consumer and Market Law 6 218.

Narayanan, Arvind. – Shmatikov, Vitaly, "Robust De-anonymization of Large Sparse Datasets," 2008 IEEE Symposium on Security and Privacy (sp 2008), Oakland, CA, USA, 2008, pp. 111-125.

Oostveen, Manon, 'Identifiability and the applicability of data protection to big data', International Data Privacy Law, Volume 6, Issue 4, November 2016, pp. 299–309.

Purtova, Nadeza, 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) 10 Law, Innovation and Technology 1.

Robertson, Viktoria H.S.E. "Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the Era of Big Data." Common market law review 57, no. 1 (2020): 161–190.

Sagiroglu, Seref – Sinanc, Duygu, "Big data: A review," 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 2013, pp. 42-47.

Santos, Cristiana – Bielova, Nataliia – Matte, Célestin, 'Are cookie banners indeed compliant with the law?', Technology and Regulation, 2020.

Sweeney, Latanya, 'Simple Demographics Often Identify People Uniquely', Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

Voigt, Paul –von dem Bussche, Axel: The EU General Data Protection Regulation (GDPR) A Practical Guide. 1st ed. 2017. Cham: Springer International Publishing, 2017.

Wasastjerna, Maria, Competition, Data and Privacy in the Digital Economy (2019), University of Helsinki.

Zalnieriute, Monika – Churches, Genna, 'When a 'Like' Is Not a 'Like': A New Fragmented Approach to Data Controllership' (2020) 83 Modern Law Review 861.

Zingales, Nicolo. 'Between a Rock and Two Hard Places: WhatsApp at the Crossroad of Competition, Data Protection and Consumer Law'. Computer Law & Security Review 33, no. 4 (August 2017): 553–558.

Primary sources

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).

Article 29 Data Protection Working Party, Guidelines on Automated decision-making and Profiling for the purposes of Regulation 2016/679. (WP 251).

Article 29 Data Protection Working Party, Guidelines on Consent under Regulation 2016/679. (WP 259 rev. 01).

Article 29 Data Protection Working Party, Guidelines on Transparency under Regulation 2016/679, Adopted on 29 November 2017. (WP 260 rev.01).

Article 29 Data Protection Working Party, Guidelines on Transparency under Regulation 2016/679. (WP 260 rev.01).

Article 29 Data Protection Working Party, Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC. (WP217).

Article 29 Data Protection Working Party, Opinion 1/2010 on the concepts of "controller" and "processor". Adopted on 16 February 2010. (WP 169).

Bundeskartellamt, Decision B6-22/16 of the FCO on 6 February 2019.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European Strategy for Data, COM(2020) 66 final, 19 February 2020.

Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules.

Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive').

European Commission, 'Mergers: Commission approves acquisition of LinkedIn by Microsoft, subject to conditions' IP/16/4284 (6 December 2016).

European Convention on Human Rights.

European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, Adopted on 4 May 2020. Version 1.1, 13 May 2020. (Guidelines 05/2020).

European Data Protection Board, Guidelines 08/2020 on the targeting of social media users, Adopted on 13 April 2021. Version 2.1, 7 July 2021. (Guidelines 08/2020).

European Data Protection Board, Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects, version 2.0, 8 October 2019. (Guidelines 2/2019).

European Data Protection Board, Opinion 5/2019 on the interplay between the ePrivacy Directive and the GDPR, in particular regarding the competence, tasks and powers of data protection authorities. Adopted on 12 March 2019. (Opinion 5/2019).

European Data Protection Board's Opinion 5/2019 on the interplay between the ePrivacy Directive and the GDPR, in particular regarding the competence, tasks and powers of data protection authorities, adopted 12 March 2019.

European Union Charter of Fundamental Rights.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM/2021/206 final.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

Online references

Alphabet, Inc. Ad Manager and Ad Exchange program policies, Tools to help publishers comply with the GDPR. https://support.google.com/admanager/answer/7666366?hl=en (Accessed 30 January 2023).

Austrian Data Protection Authority (Datenschutzbehörde) decision on March 6, 2023. https://noyb.eu/sites/default/files/2023-03/Bescheid%20redacted.pdf (Accessed 25 April 2023).

Court of Justice of the European Union, PRESS RELEASE No 91/20, Luxembourg, 16 July 2020: The Court of Justice invalidates Decision 2016/1250 on the adequacy of the protection provided by the EU-US Data Protection Shield. https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf (Accessed 31 October 2022).

Datatilsynet. Google Analytics, Recent European Practices. https://www.datatilsynet.dk/english/google-analytics (Accessed 25 April 2023).

Eggers, William D., Data as the Currency. Government's role in facilitating the exchange. 25 July 2013. https://www2.deloitte.com/us/en/insights/deloitte-review/issue-13/data-as-the-new-currency.html (Accessed 28 October 2022).

European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM (2020) 65 final.

https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (Accessed 27 April 2023).

Flender, Samuel, Data is not the new oil. Towards Data Science. 10 February 2019. https://towardsdatascience.com/data-is-not-the-new-oil-bdb31f61bc2d (Accessed 28 October 2022).

Il Garante per la protezione dei dati personali. ChatGPT: Garante privacy, limitazione provvisoria sospesa se OpenAI adotterà le misure richieste. L'Autorità ha dato tempo alla società fino al 30 aprile per mettersi in regola. https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874751 (Accessed 14 April 2023).

Il Garante per la protezione dei dati personali. Intelligenza artificiale: il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori. https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english (Accessed 14 April 2023).

Meta Reports Third Quarter 2022 Results. Meta, Inc. 16 October 2022. https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-Third-Quarter-2022-Results/default.aspx (Accessed 5 November 2022).

State of California Department of Justice. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa (Accessed 26 April 2023).

Stringer, Alyssa - Wiggers, Kyle: ChatGPT: Everything you need to know about the AI-powered chatbot. https://techcrunch.com/2023/04/25/chatgpt-everything-you-need-to-know-about-the-ai-powered-chatbot/ (Accessed 13 April 2023).

The Office of the Data Protection Ombudsman. Controller's legitimate interests. https://tietosuoja.fi/en/inform-data-subjects-about-processing (Accessed 16 February 2023).

The Office of the Data Protection Ombudsman. Inform data subjects about processing. https://tietosuoja.fi/en/inform-data-subjects-about-processing (Accessed 16 February 2023).

The Truth About Your WhatsApp Data. The New York Times. 13 January 2021.

https://www.nytimes.com/2021/01/13/technology/whatsapp-data.html (Accessed 28 October 2022).

Case law

Case C-252/21, Opinion of Advocate General Rantos delivered on 20 September 2022, *Meta Inc. et* al, ECLI:EU:C:2022:704.

Case C-252/21, Request of Oberlandesgericht Düsseldorf (Germany) for a preliminary ruling on 24 March 2021.

Case C-311/18, Judgment of the Court (Grand Chamber) of 16 July 2020, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems.* ECLI:EU:C:2020:559.

Case C-362/14, Judgment of the Court (Grand Chamber) of 6 October 2015, *Maximillian Schrems v Data Protection Commissioner*, ECLI:EU:C:2015:650.

Case C-40/17, Judgment of the Court (Second Chamber) of 29 July 2019, *Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV*, ECLI:EU:C:2019:629.

Case C-40/17, Opinion of Advocate General Bobek delivered on 19 December 2018, *Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV*, ECLI:EU:C:2019:629.

Case C-210/16, Judgment of the Court (Grand Chamber) of 5 June 2018, *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH*., ECLI:EU:C:2018:388.

Joined Cases C-92/09 and C-93/09, Judgment of the Court (Grand Chamber) of 9 November 2010, *Volker und Markus Schecke GbR* (C-92/09) and *Hartmut Eifert* (C-93/09) *v Land Hessen*, ECLI:EU:C:2010:662.

Case C-13/16, Judgment of the Court (Second Chamber) of 4 May 2017, *Valsts policijas Rīgas reģiona pārvaldes Kārtības policijas pārvalde v Rīgas pašvaldības SIA "Rīgas satiksme"*, ECLI:EU:C:2017:336.

Case C-673/17, Judgment of the Court (Grand Chamber) of 1 October 2019. *Bundesverband der Verbraucherzentralen und Verbraucherverbände - Verbraucherzentrale Bundesverband e.V. v Planet49 GmbH*., ECLI:EU:C:2019:801.

Case C-582/14, Judgment of the Court (Second Chamber) of 19 October 2016. *Patrick Breyer v Bundesrepublik Deutschland*, ECLI:EU:C:2016:779.

Abbreviations

AG	Advocate General
AI	Artificial Intelligence
CCPA	California Consumer Privacy Act
CFR	The Charter of Fundamental Rights in the European Union
CJEU	The Court of Justice of the European Union
DPA	Data Processing Agreement
ECHR	The European Convention on Human Rights
EDPB	European Data Protection Board
EEA	European Economic Area
EU	European Union
FCO	Bundeskartellamnt (German Federal Cartel Office)
GDPR	General Data Protection Regulation
GWB	Gesetz gegen Wettbewerbsbeschränkungen (Law against restrictions on competition)
LLM	Large Language Model
NYOB	European Center for Digital Rights (“none of your business")
R&D	Research and Development
SA	Supervisory Authority
USA	The United States of America
WP29	Article 29 Data Protection Working Party

1 Introduction

1.1 Background

1.1.1 Relevance and importance of the topic

The past twenty years has given a rise to new business models for companies operating online. Data has become the new oil, as it has been said.[1] Whether or not this is true[2], at least customers of some online services have become familiar with using data as a counter-performance instead of money for some services offered online. Examples of services subject to this development include, for example, social media platforms, which also have been the main target of new rules and regulations in the fields of competition law, consumer protection law, as well as data protection law. In the European Union (EU), the most significant new regulative tool in data protection law has been the General Data Protection Regulation (GDPR)[3], which has been applicable since 25 May 2018. GDPR regulates *inter alia* the processing of personal data, and right to the protection of personal data of natural persons.

User data as a payment or counter-performance or consideration[4] for the use of online services has occasionally raised concern not only among scholars and data protection professionals, but also among the society as large.[5] While data as a consideration is not a simple topic in the area of contract law[6], it unquestionably is a remarkable issue in the field of data protection law.

---

[1] The known quote is usually credited to UK mathematician Clive Humby, already back in 2006. Among others, Meglena Kuneva, then the European Consumer Commissioner, has also referred to it in her speech in 2009.

[2] There is an interesting discussion of whether data is the new oil or not. For example, Samuel Flender, a data scientist and engineer, suggests in his blog text that while being an infinite resource and being too noisy to provide value by itself, the comparison does not work very well. https://towardsdatascience.com/data-is-not-the-new-oil-bdb31f61bc2d (Accessed 28 October 2022).

[3] 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)

[4] See, for example, Eggers, Hamill, Ali, "Data as Currency" (2013) 13 Deloitte Review 18.

[5] See, for example, The Truth About Your WhatsApp Data, published in The New York Times on 13 January 2021. https://www.nytimes.com/2021/01/13/technology/whatsapp-data.html, Accessed 28 October 2022).

[6] Carmen Langhanke and Martin Schmidt-Kessel, "Consumer Data as Consideration" (2015) 4 Journal of European Consumer and Market Law 6 218.

From the customer point of view, it is understandable and even justified to collect general customer data, such as name, email and/or postal addresses, phone number, and in some cases even gender, to enable companies to offer their services to customers properly. However, there have been cases in which it has turned out that businesses have used their customer data for other purposes than just offering their products and services to customers, or they might have collected excessive information relating to their customers' behavior from third-party sources, such as other online services or data brokers, in order to produce comprehensive datasets bearing commercial value they could eventually make use of. This was the case in so called German Facebook case[7], where Facebook (nowadays known as Meta) collected (sensitive) personal information from other online services in order to produce "super profiles" of their customers, also in situations where the data subject were not even Facebook users nor aware of the fact that their personal data were being collected.

1.1.2 Fragmented regulation regarding the data collection

It is not very clear how the data collection should be regulated. Instinctively one could think that the GDPR would catch it all, but at the same time it is more complex of an issue: there are aspects of consumer (protection) law, competition law, as well as topics in a distinct and complex interplay with data protection law included.[8] The EU legislator is finding new ways to constrain increased powers of technological giants. I have limited my thesis to the assessment of GDPR, its principles, and the compliance of those business models that are reliant on user data collection. One still should not forget other legislative instruments that share interplays with data protection law and the GDPR, those being the ePrivacy Directive[9],

[7] Case C-252/21, *Meta Platforms et al.* (Opinion of the Advocate General available, case still pending in April 2023).
[8] Zingales, Nicolo. 'Between a Rock and Two Hard Places: WhatsApp at the Crossroad of Competition, Data Protection and Consumer Law'. Computer Law & Security Review 33, no. 4 (August 2017): 553–558.

[9] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications).

Unfair Commercial Practices Directive[10], and Digital Services Act[11] and Digital Markets Act[12], to name a few. Due to outer limits of this paper those legislative vehicles are mostly left out of the scope of my thesis, unless otherwise stated.

1.1.3 User data collection as a phenomenon

Collecting behavioral data of customers/users is not a new phenomenon. First attempts in trying to predict consuming preferences and behavior dates back to early 1900s, where Arthur Nielsen started to record radio, and later on television, program listening and watching statistics and to compare them with demographic information, enabling elementary profiling of the consumers of radio and television programs.[13]

Today it is possible to collect more in-depth and variable information (*data*)[14] from internet users, for example regarding their use of online products, the devices, and applications they are using the products with, their points of interests, and even location information, email addresses, social networks etc.

As a raw and unorganized data, it is not much of a value, but in hands of a capable, sophisticated, and technically competent entity, analyzed and organized data may become a highly valuable asset that may be used to research and development (R&D) internally, to enable targeted advertising, or even sold or otherwise disclosed to other parties and thus it has commercial value.

---

[10] Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council ('Unfair Commercial Practices Directive'), as amended with the Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules.

[11] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.

[12] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

[13] Julie E. Cohen, 'Everything Old is New Again – Or Is It?' in 'Between Truth and Power: The Legal Constructions of Informational Capitalism', Oxford Academy 2019, pp. 38–39.

[14] In this paper with the term 'data' I mean any kind of information having a connection to information and communications technology (ICT). This should be distinguished from the term 'personal data' which is defined later in this paper.

Some technology conglomerates today are even dependent on the user data they collect. Best examples are tech giants Meta Inc. (Facebook) and Alphabet Inc. (Google), whose revenues are generated mainly by monetizing user data through targeted advertising. Even though their business models differ, they are similar enough to fulfil my needs in this thesis as examples of data reliable businesses.

For instance, Meta Inc., formerly known as Facebook Inc., reached a revenue of 7.53 $ per user on the third quarter of financial year 2022 from all Facebook family products[15], while average revenue per user from Facebook and Messenger services only was globally 9.41 $ during the same period[16]. It is noteworthy that the USA and Canada were a lot more profitable (49.13 $ per user) geographical areas than Europe (14.23 $ per user), which is likely the consequence of stricter legislation when it comes to monetizing (personal) user data.

As advertisers are willing to pay more for efficient and targeted advertising which are more likely to increase sales in comparison to untargeted advertising, the main activity of user data reliant business is to collect user data and make it valuable through data analytics, for example by creating comprehensive user profiles to enable advertisers eventually target their advertising to those people who are most likely to use money to buy their products.

Collecting user data is necessary to the operations of such business models, as they monetize user data through targeted advertising. Meta tells in its 2021 Annual Report that reducing the amount and quality of user data could lead to a situation where they must cease their operations in the EU.[17] Furthermore, those business models must modify their privacy policies as well as terms of service in order to comply with changing legislation and other data protection law requirements, which could harm their ability to gain revenue. Meta reveals the GDPR and ePrivacy Directive as laws that influence them the most, and more specifically the GDPR have impacted their ability to use data signals from user activity on websites and services that they do not control, and moreover, increasing number of users have opted to control their data use after the adoption of the GDPR.[18]

---

[15] Meta Earnings Presentation Q3 2022, slide 12.

[16] Meta Earnings Presentation Q3 2022, slide 15.

17 Meta Annual Report 2021.

18 Meta Annual Report, Item IA. Risk Factors, p. 17.

Another challenge Meta has faced is the invalidation of so-called Privacy Shield, an agreement between the EU and the United States of transfer framework for data being transferred from the European Union to the United States, which was invalidated by the Court of Justice of the European Union (CJEU) in July 2020. Meta refers to Schrems II case (Case C-311/18, *Data Protection Commissioner v. Facebook Ireland and Maximilian Schrems*), in which the CJEU invalidated Decision 2016/1250 on the adequacy of the protection provided by the EU-US Data Protection Shield. In its ruling, the CJEU held that the Privacy Shield was invalid, but the Standard Contractual Clauses (SCCs) still is a valid transfer framework for data transfers between the EU and the US[19]. Due to limits of the paper, I must, however, leave the data transfer regime out of the scope of my research.

1.2 Research question, methods and materials

1.2.1 Research questions and focus of the thesis

Against the background provided so far, the main research question this paper intends to investigate is:

> Under which conditions may personal data be collected through sources other than the data subject under the GDPR?

To be able to answer the question we must answer the following questions as well:

> What is considered as personal data, and what is considered as sensitive personal data?
>
> Who are the stakeholders in data collection system and what are their liabilities towards users and their rights as meant in the GDPR?
>
> Which party is liable towards users to ensure users' rights in cases where the data is not collected directly from the users?
>
> How do the principles of data minimization and purpose limitation limit the collection of personal data from third parties?

In this paper I am going to research the conditions under which one is or is not allowed to collect personal data from other parties than the data subject under the GDPR. The subject

---

19 See, in brief, Press release of the Court of Justice of the European Union regarding the Judgment in Case C-311/18, 16 July 2020, available at https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf. Accessed 31 October 2022.

matter is quite topical and the significance of it will only increase with accelerated tempo of the development of the artificial intelligence (AI) and Big Data, and the increased quantity of available (personal) data, because the access to data is important in the development process of AI. Usually, the datasets contain or at least has contained personal data, and in those situations the data have been collected in many cases from sources other than the data subjects themselves.

Firstly, I provide a general background of the GDPR, in relevant parts to provide a theoretical background of the subject. Secondly, I specify key stakeholders, both in practice as well as in GDPR terms. Thirdly, I go through information obligations towards the data subjects of those stakeholders. Fourthly, I separate three different legal bases under which user data is collected, and who is liable for enhancing the use of user rights as meant in the GDPR, and what conditions apply to data processing. Those legal bases are also applied to the AI development purposes. Next, a look at the German Facebook case, which is still an ongoing case in the Court of Justice of the European Union, is taken to illustrate the collection of personal data from other sources in real life. Lastly, in addition to conclusions a brief oversight of the future of the data protection law is discussed.

Key findings and answers to research questions may be summarized as follows:

When collecting data from sources other than the data subjects themselves, controllers must bear in mind the informational obligations under Articles 12-14 of the GDPR as well as make sure they can legitimately rely on one of the legal bases under Article 6(1) for data processing, of which the collection of data is just one example. Consent as a legal basis could provide most trustworthy legal basis, if and when the consent has been obtained efficiently. Obtaining the consent when the source of the personal data has been other than the data subject may sometimes be challenging, but it is even possible for other independent or joint controllers to rely on a consent that has been originally obtained from the data subject, as long as they have been named and the data subject has provided all relevant information on the purposes of the processing as well as all organizations that are taking part in the processing.

Performance of a contract may in some limited occasions be relied on when collecting personal data from third-party sources. In practice the use cases are limited to the steps taking place before entering into a contract with the data subject, at the request of the data subject,

such as to check address details from public registers, as long as the processing is necessary for the performance of the contract.

A third legal basis that could be relied upon is the legitimate interests of the controller or of a third party, which contributes certain flexibility to the GDPR. However, controllers are obliged to plan the processing activities and legal bases beforehand, and they should not rely on the legitimate interests as a last resort in case that other legal bases stop functioning to their needs. What is more, the fundamental rights and freedoms and reasonable expectations of the data subjects should be taken into account, and to not override them.

Whichever is the legal basis that the controller relies upon, the principles of the data minimization as well as purpose limitation limit and control the amount, nature, and exploitation of the personal data being collected. Purposes of the processing should be defined beforehand, and the data should not be used to purposes that are contrary to the original purposes. The amount and nature of the data should also be limited to the minimum that is needed in order to fulfil the processing purposes, given that the processing is otherwise lawful as well.

In the end, each controller bears the burden of proof for the compliance with the GDPR at any given time, which forespeaks to the proactive planning and organizing of the data processing activities.

Even though the GDPR has set an example in the data protection law globally, it has not been free of criticism, especially in the rise of artificial intelligence. Strict regulations and obligations set forth may hinder technological development and innovation. From the point of view of the data subject the GDPR still provides protection. More discussion, both among the experts and the public, is needed in reconciling the innovation and right to privacy in accelerating development of datacentric technologies.

1.2.2 Methodology and materials

The thesis paper is a legal overview of data protection law, also drawing on texts related to competition and consumer law, focusing naturally in the requirements and obligations stemming from the GDPR to different parties taking part in the data collection framework. The aim of the case studies is to provide the reader with practical examples of situations in which the collection of data from third parties has proved problematic in some way.

From competition law point of view there has been a quite lot research regarding data collection and use, for example Viktoria H.S.E. Robertson's article on excessive data collection[20], in which she concludes that competition law as such provides necessary tools to address issues arising from excessive data collection.

The thesis of the article is that competition law in the EU already possesses tools to include privacy considerations into competition law analyses. While describing relevant market for data collection, she describes the relationship between data collector and user being comparable to traditional consumer relationship, but with the difference that in online services users "pay" for the service with their personal data, instead of money.[21] It is relevant to understand the whole picture with other stakeholders as well, and to recognize their roles also under the terms of the GDPR.

When it comes to competition law analysis, according to Robertson, breach of data protection rules is not automatically a breach of competition law, but rather data protection law can provide depth to competition law analyses.[22] Excessive data collection and its implications on competition law is at the heart of the article. In the area of competition law, privacy concerns are relevant in assessing abuse of dominance or controlling mergers. What is more, European Commission has stated privacy issues relevant to competition law analysis in its investigation of LinkedIn's and Microsoft's merger.[23] Also, European Data Protection Board held that assessing privacy issues is essential in the assessment of potential abuse of dominance and in merger controls in its statement regarding Apple/Shazam merger.

Maria Wasastjerna has written a doctoral thesis[24] of the interplay between data protection regime and competition law, and asks what kind of role does, and should, personal data and privacy have in competition law. The study also provides a good framework for assessment of online business models that rely on user data. Even though more in-depth analysis on

---

[20] Robertson, Viktoria H.S.E. "Excessive Data Collection: Privacy Considerations and Abuse of Dominance in the Era of Big Data." Common market law review 57, no. 1 (2020): 161–190.

[21] Robertson 2020, p. 170.

[22] Robertson 2020, p. 168.

[23] European Commission, "Mergers: Commission approves acquisition of LinkedIn by Microsoft, subject to conditions" IP/16/4284 (6 December 2016).

[24] Maria Wasastjerna, Competition, Data and Privacy in the Digital Economy (2019), University of Helsinki.

competition law matters is left out from my thesis, it provides as a good analysis of changing online market and data collection as a phenomenon.

Research regarding the principles of data minimization and the purpose limitation under the GDPR and its interpretation to data collection through third parties is scarce. The regulation is still quite young, and case law still developing. To my purposes, it has been fruitful to read companies' annual reports and their challenges regarding compliance with the GDPR and other legislative requirements, such as ePrivacy Directive.

In understanding the GDPR itself, great handbooks have helped me: "Uusi tietosuojalainsäädäntö"[25] and "The EU General Data Protection Regulation (GDPR) : A Practical Guide"[26]. Those handbooks have provided a lot of interesting references to dig deeper into the world of data collection business.

Keeping the focus in data collection from other sources than the data subjects themselves has proofed to be somehow challenging due to complexity of the legislation and the differences between real-life practices and the theoretical framework. To get back in track some sources outside the most usual sources in legal studies consist of annual reports and interim reports, which are great sources of information to understand challenges regarding compliance with data collection requirements, among others. Especially, I focus on the leading social network platform, Meta, Inc, formerly known as Facebook, Inc, which is a parent company to a bunch of widely used social network platforms, such as Facebook, Instagram, and WhatsApp. Other relevant user data reliant online business is, for example, Alphabet, Inc, which is the parent company of Google, YouTube, and Android, to name a few.

Case law and the consequences it has had in the user data reliant businesses has a significant role in my thesis. Most importantly, the *German Facebook case* (Case C-252/21 (only AG opinion available as of 31 October 2022)) is assessed among ohters, not to forget the *Fashion ID* case.

[25] Korpisaari – Pitkänen – Warma-Lehtinen 2008.

[26] Voigt, Paul., and Axel. von dem Bussche. The EU General Data Protection Regulation (GDPR) A Practical Guide. 1st ed. 2017. Cham: Springer International Publishing, 2017.

1.2.3 Limitations of the thesis

There are some issues that must be left out of the research due to limits of the paper. Practically the most significant of those in the field of use and reuse of user data is the data transfer regime, because largest technological giants are mostly located in the United States, while the GDPR protects the users located in Europe. In practice, where the data is being transferred from the EU/EEA area to third countries, such as the United States, the data transfer must be legitimated under the obligations of the GDPR and case law of the CJEU.[27]

Another topic sharing a close linkage to the GDPR and the data collection is the ePrivacy Directive, also known as the "cookie law". The ePricacy Directive is only referred to briefly, as in comparison to the GDPR it could be seen to complement the GDPR.[28]

As a juridical analysis on the data protection law this paper does not aim to provide technical solutions for compliance with the GDPR. However, some sources for more information on technical requirements are provided along with the legal obligations and requirements.

Competition law, consumer law, and fundamental rights and freedoms are assessed only insofar as it has been necessary to understand the context. The focus has been kept in the data protection law.

[27] See more of the data transfers to third countries under the GDPR after *Schrems II* decision in Corrales Compagnucci, Marcelo – Mateo, Aboy – Minssen, Timo, 'Cross-Border Transfers of Personal Data After Schrems II: Supplementary Measures and New Standard Contractual Clauses (SCCs)', 30 December 2021.

[28] Article 1 of the ePrivacy Directive; Article 94(2) of the GDPR. See more of the functioning of the ePrivacy Directive with the GDPR on European Data Protection Board's Opinion 5/2019 on the interplay between the ePrivacy Directive and the GDPR, in particular regarding the competence, tasks and powers of data protection authorities, adopted 12 March 2019.

2 Data collection and the GDPR

2.1 Scope of application of the GDPR

The scope of application of the GDPR is wide and comprehensive to ensure a high-level protection of personal data.[29] According to the Article 2(1) of the GDPR, the GDPR applies to any wholly or partly automated processing of personal data, as well as processing of personal data without automated means, if personal data is forming or is meant to form part of a filing system. Against the background of my research, the collection of user data is in practice always under the scope of application of the GDPR. The GDPR applies also to monitoring of behavior of natural persons online, for example by using or reusing personal data to analyze his or her choices or preferences or in order to create a profile of him/her.[30] Targeting or personalizing advertisements could be considered to meet the definition of behavioral monitoring.[31]

Geographically, the GDPR applies not only when establishments process personal data within the EU, but also where the establishment located outside the EU is processing the personal data of individuals located in the EU.[32]

2.2 Relevant definitions under the GDPR

2.2.1 Personal data

Since the GDPR applies to the processing of personal data, it must be further defined what is personal data. According to the definition provided in the Article 4, Section 1 of the GDPR, personal data "means any information relating to an identified or identifiable natural person". Identifiability is in turn defined as a possibility to identify a person "directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". Therefore, data is

---

[29] Voigt – von dem Bussche 2017 p. 9.

[30] Korpisaari – Pitkänen – Warma-Lehtinen 2018, p. 48.

[31] Lukas Feiler, Nikolaus Forgó, Michaela Weigl, 'The EU General Data Protection Regulation (GDPR): A Commentary. Globe Law and Business 2018, p. 54.

[32] Article 2 of the GDPR; Voigt – von dem Bussche 2017, p. 22.

personal data if it may be connected to any natural person and identify his or her. It is not clear from the wording of the Article 4(1) of the GDPR that is it the data controller or data processor (as defined in the GDPR and later in this paper) that should possess the additional information in order to identify the natural person in question (data subject, as defined later), or could it be anyone. Voigt interpretates it that the wording suggests it could be anyone.[33] On the other hand, it is ultimately a question of how likely and by what means the controller can identify the person.

To illustrate the importance of identifiability and consequently defining what is personal data, we could imagine a situation where internet-browsing behavioral data is collected from a visitor (internet user) at the webpage of a local newspaper, where the data consist of the time spent per article, clicks on the webpage, information on the gadget used to visit the webpage, as well as information on the other open or recent tabs on the web browser. Without further information this is most likely not considered as personal data, since the webpage visitor is not likely to be identifiable from this kind of information. However, if the data is collected by a social media platform, which already possess personal data of the particular visitor of the webpage in question, in hands of it the data becomes personal data if the collected data is relating to the webpage visitor. However, it is not required that the information is in the possession of one actor, but it is enough if the additional information is available to the controller/processor in question by reasonable likely means.[34]

2.2.2 Sensitive personal data

Paragraph 1 of Article 9 of the GDPR defines ‘special categories of personal data’ (usually referred as ‘sensitive personal data’) as personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership; genetic or biometric data when used to identify a natural person; and data concerning health or sexual orientation. In short, these kinds of data have stricter protection regime as specified in paragraph 2 of Article 9 due to the personal and sensitive nature of them. Paragraph 4 allows Member States to possess further conditions regarding the processing of genetic or biometric data or data concerning health. This additional layer of protection regarding sensitive data includes, among others, the explicit consent regime as meant in subparagraph (2)(a) of the

---

[33] Voigt – von dem Bussche 2017, p. 11.

[34] Korpisaari – Pitkänen – Warma-Lehtinen 2018, p. 58.

Article 9. Protection of sensitive personal data is further assessed in chapter regarding the German Facebook case.

2.2.3 Data controller, data processor, and joint controllers

Definitions of data controller, processor as well as joint controllership is of great significance in this paper. As they may seem simple at first glance, distinguishing them from each other may be a more difficult exercise. Article 4 of the GDPR defines 'controller' as a natural or legal person which determines the purposes and means of processing of personal data, and 'processor' as a natural or legal person which processes personal data on behalf of the controller.

In intra-group relations as well as in cooperation relations between businesses it is not always clear who is the controller, who is the processor, or are two or more businesses joint controllers. Some characteristics of the controller are that the entity in question is free of guidance or control of other entities when processing personal data, it can use the data for own purposes instead of purposes of the other entity, and it is responsible of the processing for the lawfulness and accuracy of the data processing towards both the supervisory authority and data subjects. Thus, where an organization is using a processing service such as data storage services in cloud, the organization is usually the controller and the service provider of the cloud service a processor.[35]

To "determine the purposes and means of processing" does not cover every single decision, but instead it covers most important guidelines and crucial parts of the processing, such as, but not limited to, why is the data processing taking place, for how long, which data are collected, and for how long are they stored.[36] Nonetheless, determining the purposes of the processing is not just a theoretical framework, but instead an obligation of the controller, as the data subjects should be informed of those purposes as well under the informational obligations under Articles 12-14 of the GDPR.

---

[35] Voigt – von dem Bussche 2017, p. 18–19.

[36] Voigt – von dem Bussche 2017, p. 19; Art. 29 Data Protection Working Party, WP 169 (2010), p. 8.

Processor, on the other hand, is the subcontractor of the controller, who is lacking the decision-making power over the data processing and is a separate entity with respect to the controller.[37] Without the controller there is no processor either.

Where some of the decision-making powers are allocated to both the controller and its business partner, instead of forming a controller-processor relationship, it could constitute a joint controllership. According to Article 26(1) of the GDPR, "[W]here two or more controllers jointly determine the purposes and means of processing", they form a joint controllership, and the parties are joint controllers. Joint controllers are both responsible towards data subjects for the data processing (Article 26(3) of the GDPR). Joint controllers shall allocate the responsibilities of the data processing together in a transparent way.

Joint controllership as a concept has not been free of confusion. In the criticized *Fashion ID* case[38] the CJEU expanded the definition of the concept of joint controllers to cover relationships where the responsibilities were not allocated between the parties. In its decision the CJEU held that the *Fashion ID*, a webpage operator and an online store, was deemed to have a joint controllership with Facebook (today: Meta, Inc.) up until the data is being subsequently processed by Facebook, and from there on Facebook is the only controller.

However, it must be noted that the case at hands was referring to the EU Directive 95/46 (Data Protection Directive), the predecessor of the GDPR, but it was delivered on 29 July 2019 by the CJEU, that is, when the GDPR had already entered into force and the application of it had begun. The Directive was applicable in the *Fashion ID* case since the proceedings began in 2015.[39]

In addition to the *Fashion ID* decision the CJEU has shaped the definition of joint controllers in the *Wirtschaftsakademie* decision, where it held that the administrator of a "fan page" on

---

[37] Voigt – von dem Bussche 2017, p. 20, Art. 29 Data Protection Working Party, WP 169 (2010), p. 25.

[38] Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV ECLI:EU:C:2019:629 (Fashion ID).

[39] For more comprehensive analysis on joint controllership and the *Fashion ID* case, see Monika Zalnieriute and Genna Churches, 'When a 'Like' Is Not a 'Like': A New Fragmented Approach to Data Controllership' (2020) 83 Modern Law Review 861.

Facebook is also determining the purposes and means of the processing together with Facebook and thus shall be regarded as a joint controller.[40]

2.2.4 Data subject, third parties, recipients of data

The data subject, recipients of data, and third parties are also among the relevant parties in the data collection framework. The data subject, as the name refers to, is a natural, identified or identifiable person whose personal data are being collected [41].

Third party is defined in the GDPR as "a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data"[42]. The definition is important when assessing the legitimate interests of the controller *or third parties* under the legal basis of pursuing legitimate interests in the Article 6(1)(f) of the GDPR. The term should be distinguished from the parties other than the data subject, which I use in my research questions. The latter may be, in this paper, an independent controller, a joint controller, a processor, or a *third party*, who provide the controller the personal data of the data subject. For the sake of clarity, in this paper I use terms *data collector* and *data recipient*.

Data recipients are those entities to which the personal data are being disclosed, whether a third party or not. Recipients may be natural or legal persons, or even public authorities, unless they receive personal data in the framework of a particular inquiry.[43] In practice those data recipients often are independent controllers, but sometimes processors or even joint controllers, as will be seen in this paper.

---

[40] CJEU, Judgment in Wirtschaftsakademie, 5 June 2018, C-210/16, ECLI:EU:C:2018:388, paragraphs 36–39.

[41] Article 4 item 1 of the GDPR.

[42] Article 4 item 10.

[43] Article 4 item 9.

2.3 Lawfulness of data processing – principles of the GDPR

2.3.1 Principle of data minimization

In order to comply with the GDPR, controllers must be able to demonstrate their compliance at any given time. This has established an ex-ante regulatory framework of data protection law in the European Union.

One of the foundational principles of the GDPR is the principle of data minimization, which is the backbone of the data protection system. According to Article 5(1)(c) of the GDPR, the personal data which will be collected shall be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed". Recital 39 further specifies that the data minimization principle requires storing periods to be set to a strict minimum and the controllers to disclose the time limits for erasure of the personal data.

The principle of data minimization does not require the controllers to limit their processing activities to the absolute minimum, but rather to minimize the volume of the data collection.[44] In principle, the data minimization principle affects and limits the business activities of the data collector by setting rules to the collection and requiring data controllers to specify the purposes of the data processing and to evaluate those against the intended data collection activities: each collected datapoint should fulfil the purposes of the data processing.

2.3.2 Purpose limitation principle

The purpose limitation principle requires the personal data being collected only for specified, explicit and legitimate purposes, which shall be specified in advance, and not to further process contradictory to those purposes.[45] By requiring the personal data being processed only insofar as the purposes of the processing could not be fulfilled by other means, the purpose limitation principle is closely tied to the data minimization principle.

The importance of the purpose limitation principle lies in keeping the data subjects informed of the processing purposes and activities of the controller, and hindering the controllers to utilize the personal data to other purposes than it has been disclosed to the data subjects.

---

[44] Voigt – von dem Bussche 2017 p. 90.

[45] Article 5(1)(b); Recital 39 of the GDPR.

Indeed, collecting extensive datasets consisting of large amounts of personal data should not occur without detailed information for data subjects on how the data are going to be processed and to which purposes.[46]

[46] For more detailed analysis on principles of data minimization and purpose limitation, see Finck, Michele – Asia J. Biega, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems'. Technology and Regulation, 7 December 2021, pp. 44–61.

3 Information obligations for data collector and data recipient

3.1 Set-up

This section aims to outline the relevant GDPR obligations for data collectors and data recipients to inform the data subjects of their rights as well as the means and purposes of the processing in three different scenarios, in which personal data could be collected from parties other than the data subjects themselves. The set-up to three different scenarios is artificial and is mostly based on the GDPR vocabulary and functioning. In real life collection of personal data may take several kinds of forms, and it is even possible that different stakeholders may not strictly comply with the obligations stemming from the GDPR towards data subjects and therefore do not enable data subjects to enforce their rights, or that in upcoming legal proceedings those roles are changed or interpreted differently, as will be seen in the case *Fashion ID* (defined and assessed later in this paper).

In the first scenario, the entity which collects the personal data of the data subject (data collector) and is about to transfer the data to another controller (data recipient), is an independent data controller, which presupposes that the controller has *de facto* independent power and ability to decide on the purposes and means of its data processing. According to Article 4(7) of the GDPR, a controller is an entity which alone (or jointly with others) determines the purposes and means of the processing of personal data.

In the second scenario, the data collector is not an independent controller, but performs services for the controller whereby personal data is being collected and/or analyzed, and consequently transfers them to the controller according to their contract or other arrangement, usually known as data processing agreement (DPA). The party working for the controller is called a processor, and it presupposes that the processor does not have its own independent authority to decide on the purposes and means of the data processing.

In the third scenario those two market players, data recipient and data controller, form a joint controllership, which means that they are both counted as joint controllers, and the processing is determined and organized by those entities together, and therefore the processing purposes, means and the depth of their cooperation, legally and commercially, may vary a lot in real life. The concept of joint controllership has been subject to change in the case law of the CJEU, and in real life the roles might not always be that clear in the beginning when establishing business relations and the means of data processing.

3.2 Information obligations of the data collector under the GDPR when collecting personal data from the data subject

3.2.1 Real-life data collection examples

In order to provide a description as accurate and complete as possible of the research question, it is essential to briefly describe the information requirements applicable when data, in a first step and before being further transferred, is collected from the data subject. First, we are going to assess the obligations of the data collector. For example, data collection could happen when a service provider collects personal data of the user including but not limited to behavioral browsing habits (browsing history, other open tabs, etc.), technical information of the devices used, location and the contents the user has created, and then sells the collected and sometimes further analyzed data to its own business partners, given that those organizations are not determining the purposes of processing together.

One may think of Google (Alphabet Inc.) collecting various datapoints of its customers when they are using the Google search engine, analyzing the data, and matching them with the Google user profiles, and then put the information including personal data sale for online advertisers to enable them to show personalized advertisement to Google users. On its support webpage for advertisers, Google clarifies that the advertisers using their ads services (Google Ad Manager, Ad Exchange, AdMob, and AdSense) are independent controllers of personal data under the GDPR framework.[47] However, one should note that not all the data that is being shared to other independent controllers are personal data, and what is more, one should take a critical standpoint when analyzing the roles: the CJEU or even supervisory authorities may interpret their roles and obligations differently, for example to establish a *de facto* joint controllership between those market players, as was the case in *Fashion ID* case.

As mentioned earlier, data is *personal data* when the information relates to an identified or identifiable natural person. In presented scenario, mere technical information of the devices used to access Google services is unlikely to be labelled as personal data. Nonetheless, if the recipient of the personal data, that is the advertiser in this case, already possesses personal data of the data subject and the data is merged with previously possessed personal data, the

[47] Ad Manager and Ad Exchange program policies, Tools to help publishers comply with the GDPR, available at https://support.google.com/admanager/answer/7666366?hl=en. Accessed January 30, 2023.

newly received data is likely to become personal data as well.[48] In the same way the data in question is personal data if possessed by Google and bundled with other personal data relating to the data subject.

Tracking pixels are sophisticated tools to harvest user data from webpages. Tracking pixels, a type of "cookies", consist of small snippets of code downloaded into the computers or other gadgets of the user. Tracking pixels may monitor the behavior of the user and save information relating to open tabs, ads he or she has already seen, time spent on webpages or articles, and so on. The collected data may be used to put the individual data subject to target audiences and to create user profiles consisting of characteristics of individuals to better target advertising to them.[49] The ePrivacy Directive is an integral part of the tracking pixel legislative framework, and due to ePrivacy Directive being outside the scope of this paper, the legislative and technical requirements concerning the cookie-consent framework is not assessed further here.[50]

Two of the most widely used tracking pixels are Google Analytics and Meta Tracking Pixel. The former has faced a strong headwind in the EU followed by 101 complaints by the organization None of Your Business (NYOB).[51] Meta has faced setbacks as well, by the same organization, which lodged a complaint to the Austrian Data Protection Authority, among others. The decision found the Meta Tracking Pixel violating the GDPR.[52]

3.2.2 Data collector as an independent controller

In the first scenario the two market players (data collector; data recipient) that are taking part in the data processing are independent from each other and determine their purposes and

---

[48] The difference of these two situations could be illustrated as follows: if the advertiser receives data disclosing technical details of all the gadgets that had been used to receive their advertisement, and the data it is receiving does not include any other data, it is most likely not personal data. However, if the advertiser could link the "eyeballs" (persons seeing their ads online) with their gadgets, the data should be counted as personal data.

[49] EDPB Guidelines 08/2020 Footnote 69 p. 21.

[50] See a well-reasoned and explained study of the legal-technical requirements of the tracking pixels in Cristiana Santos, Nataliia Bielova and Célestin Matte, Are cookie banners indeed compliant with the law?, Technology and Regulation, 2020, 91–135.

[51] See an oversight of the decisions in the webpage of Danish Data Protection Supervisory *Datatilsynet* at https://www.datatilsynet.dk/english/google-analytics (accessed April 25, 2023).

[52] Austrian Data Protection Authority's (*Datenschutzbehörde*) decision on March 6, 2023, available at https://noyb.eu/sites/default/files/2023-03/Bescheid%20redacted.pdf (in German. Accessed April 25, 2023).

means of their processing activities independently. As an independent controller the entity that is primarily collecting the data from the data subject and subsequently transferring the data to other entities (data collector) must comply with all the obligations that are laid to controllers under the GDPR. Since this paper is not a comprehensive guide to understanding the GDPR, I focus only on the specific obligations of the entity which aims to first collect and subsequently transfer the personal data to other entities. It is worth mentioning that in this paper with transferring the data I mean all kinds of data transfers, but I leave the regulation of data transfers to third countries or international organizations as meant in Chapter 5 of the GDPR out of the scope of this paper due to research economics. Still, it is a much relevant aspect of the data collection, usage, and reusage in real life, and would require further investigation, as the data transfer rules especially between the EU and the US are rapidly changing due to market relevance and the changing legal proceedings and transatlantic agreements.[53]

In addition to other requirements of the GDPR and more specifically of complying with the principles relating to processing of personal data (Article 5), the lawfulness of data processing (Article 6), and the responsibilities of the controller as laid down in the Article 24 of the GDPR, a data collector must comply with the requirements under Articles 12, 13, and 14 of the GDPR, regarding informing obligations towards the data subject.

According to the Article 12 of the GDPR, the controller must inform the data subject of the information specified in Articles 13 and 14 as well as the rights of data subject and the enforcement of them to the data subject (rights of the data subject are covered in Articles 15-22 and 34 of the GDPR) in a concise, transparent, intelligible, and easily accessible form, using clear and plain language. Failure to comply with the Articles 12-22 may lead to, according to the Article 83(5) of the GDPR, an administrative fine of up to 20 MEUR or in case of an undertaking up to 4 % of the total worldwide annual turnover of preceding financial year, whichever is higher. Information obligation of the controller is the main vehicle to enforce the principle of transparency as meant in Article 5(1)(a).

The information obligation is not unlimited. Article 11 of the GDPR states that the controller is not required to maintain, acquire, or process additional information in order to identify the

53 Most notable recent case law that has been shaped the data transfers regime has been the Schrems II decision, which has been the backbone of the recent decisions banning the use of tracking pixels in the EU.

data subject only in order to comply with the GDPR. Where the controller can demonstrate that it cannot identify the data subject, it does not need to provide the information to the data subject, unless the data subject provides the controller such information that makes the data subject identifiable and will exercise his or her data subjects' rights. In this case the controller cannot deny receiving such information only to avoid its GDPR related obligations unless the controller can prove that it still cannot identify the data subject.[54]

The obligation to provide information regarding the data processing as well as the rights of data subject should be distinguished from the obligation to maintain a Record of Processing Activities as meant in the Article 30 of the GDPR, but at least the Finnish Data Protection Authority (in Finnish: *Tietosuojavaltuutetun toimisto*) has held in its guidelines that the Record of Processing Activities could be used in planning the means of information to be provided under Article 12.[55] For the sake of clarity, the Record of Processing Activities under Article 30 is an internal document (and not assessed in this paper further), whereas the fulfilment of the information obligation under Articles 12-14 is not tied to an explicit, specifically named document nor other strict formal requirements.[56] Most common practices to fulfil information obligations appears to be to inform data subjects through "Privacy notice", "Privacy policy", or "Privacy" documents/statements.

An interesting situation would actualize in a case where the data collector has collected user data and cannot identify the data subject from the data and transfers the data to a larger platform that already possesses personal data (i.e., the receiving party can identify the data subject from the data received) of the data subject. The data collector, supposing it can rely on Article 11 exception, has no information obligation towards the data subject. This could lead to a loophole in data privacy from the point of view of those collectors, if some data collectors are allowed to collect data without the need to comply with GDPR requirements, while they may provide large quantities of personal data to other entities where the same data, that has originally been non-personal data, may turn into identifiable personal data in the hands of the recipient. However, those entities which receive that kind of user data from data collectors are not excluded from the scope of the GDPR as long as they process personal data, that is

---

[54] Article 12(2) GDPR.

[55] https://tietosuoja.fi/en/inform-data-subjects-about-processing (accessed January 31, 2023).

[56] Certain requirements still apply, such as it must be made available upon collection or obtainment of the personal data and in an easily accessible form. See, for example, Voigt – von dem Bussche 2017 p. 143.

identifiable data that could be linked to natural persons (data subjects). I will study that situation in the next subchapter from the point of view of the data recipient.

Keeping in mind the purpose of this paper, the most relevant informational obligation of data collector stems from the Article 13 which obliges the controller to provide data subject certain information where the personal data are collected from the data subject itself.[57] According to the Article 13(1), at the time when personal data are obtained, the controller shall provide the data subject information on the controller, its possible representative and/or data protection officer, the purposes and the legal basis of the processing including the legitimate interests of the controller (Article 6(1)(f) of the GDPR), the recipients or categories of recipients of the personal data as well as certain details on data transfers to third countries or international organizations, if applicable.

Pursuant to Article 13(2), the controller must provide additional information on the period or criteria of period for which the personal data will be stored, how to enforce data subject's rights, revoking of the consent, lodging a complaint with a supervisory authority, whether the provision of personal data is a statutory or contractual requirement or a requirement necessary to enter into a contract, as well as whether the data subject is obliged to provide the personal data and of the possible consequences of failure to provide such data, and the existence of automated decision-making, including profiling, if applicable.

Additionally, paragraph (3) requires the controller to provide all relevant information on changed purposes of processing to the data subject prior to such processing. One still should keep in mind that even though the controller aims to use personal data to other purposes, it should take into account the principle of purpose limitation (Art. 5(1)(b)) as well as restrictions stemming from the Article 6(4) regarding the context and nature of such personal data, and possible links between the original and changed purpose.

Finally, paragraph (4) states that the information obligation does not apply insofar as the data subject already has an access to or otherwise possesses the information.

Privacy notice or equivalent vehicle of information should be kept up to date, and the data subjects informed of any substantial and material changes to it. Substantial and material

---

[57] Article 14 of the GDPR, on the other hand, requires the controller to provide almost identical information when the source of the personal data has been other sources than the data subject. This situation is assessed in the following subchapter.

changes may be, *inter alia*, changed or modified i) processing purpose, ii) controller or iii) means to enforce the data subjects' rights.[58] Substantial and material changes to the policies must be informed in noticeable and timely manner.[59]

Given the subject matter of this paper, the most relevant part of Article 13 is the obligation for the controller to inform the data subjects of the recipients or categories of recipients of the personal data.[60] Therefore, data collector should not collect personal data and transfer them to other recipients than what has been disclosed to data subjects without first disclosing the new recipients to data subjects. The information to be provided to data subjects of data collection including the recipients of personal data must not be buried between the lines of privacy notices among other information. Article 29 Data Protection Working Party (hereinafter "WP29") has recommended a layered approach to be used to inform data subjects, according to which privacy notices should be categorized and linked in a layered manner rather than to state all privacy statements/policies in a single document.[61] Data subjects should be enabled to get a clear overview of the policy and dig deeper one category at a time.[62]

Moreover, data subjects should be informed where personal data can be legitimately disclosed to other recipients and when the personal data are first disclosed to said recipient(s).[63] Thus, without the knowledge of the data subject, personal data should not be transferred to other recipients. Where data collector has not specified the data recipients or categories of recipients of data, the privacy notice should be amended and the material and substantial changes to it informed to the data subjects. Additionally, when personal data is lawfully disclosed to recipients of data for the first time, data subjects should be informed, and should

---

[58] WP 260 rev.01., pp. 16–17.

[59] *Ibid.*

[60] Article 13(1)(e) of the GDPR.

[61] WP 260 rev.01., p 19.

[62] However, it is not always true that layering and categorizing the information in a webpage would be the clearest way for data subjects to access the relevant information. Contrary to what the WP29 has proposed, some privacy experts prefer a single document which would enable search features best, that is Ctrl+F.

[63] Recital 61 of the GDPR.

the purpose of the data processing be other than for what were the data originally collected, shall the controller inform the data subject of those changes and other relevant information.[64]

3.2.3 Data collector as a processor

In this second scenario the data collector acts as a processor, which works for the controller according to their contractual arrangements. Obligations of the processor are much lighter than those of the controller. The GDPR obliges the processor to enter into a written contract (Data Processing Agreement) with the controller, which is determining the purposes and means of the processing.[65] Where the processor steps in and starts to determine the purposes and means of the processing independently, it shall be counted as a controller.[66]

The processor is first and foremost accountable for its processing activities towards the controller, subject to their contract or other legal act, although there is an indirect responsibility to assist the controller to comply with the obligations to respond to requests for exercising the data subject's rights by appropriate technical and organizational measures, taking into account the nature of the processing as well as the contract between the controller and the processor.[67] Other obligations of the processor stemming from Article 28 are more technical of their nature.

The processor is not liable towards the data subjects more than to assist the controller with the requests to enforce data subjects' rights. Therefore, the processor should be seen as a subcontractor for the controller in the sense of the GDPR, even though real-life situations may vary.

3.2.4 Joint controllers

As it has already stated earlier, joint controllers are together determining the purposes and means of the processing, and the roles are not always very clear, not even for the parties themselves. Join controllers taking part in data collection and further processing arrangements are determining the purposes and means of the processing together, and they should make it

---

[64] Ibid.

[65] Article 28(3), (9) of the GDPR.

[66] Article 28(10) of the GDPR.

[67] Article 28(3)(e) of the GDPR.

available to the data subject as well.[68] Nonetheless, according to Article 26(3) of the GDPR, the data subject may exercise his or her rights in relation to either of the two (or more) joint controllers.

As joint controllers are controllers in the same way as ordinary controllers, the same obligations and responsibilities that has already been assessed apply to them as well, unless they have transparently and in a clear manner divided such obligations between themselves and informed the data subject thereof. This should not cause any trouble as far as those arrangements are clear for each party taking part in the data collection. If that is not the case, as it turned out in the *Fashion ID* with the CJEU's "fragmented approach to joint data controllership"[69], the responsibilities may become fragmented as well.

In the *Fashion ID* case the fashion retailer Fashion ID collected personal data of its webpage visitors by a social plug-in, a Like-button, and Facebook could access the data consisting of, among others, IP addresses, even when visitors were not Facebook users.[70] Fashion ID allegedly failed to inform their webpage visitors of the arrangement as it should have been according to the Data Privacy Directive, which was in force then.[71] Fashion ID claimed that they did not even have an access to such data, nor were they aware of the data collection.[72] As it has been already mentioned, the CJEU found the parties to form a joint controllership, limiting the Fashion ID's controllership to the collection and disclosure of the personal data to Facebook.[73]

CJEU was of an opinion that Fashion ID was aware of the data collection, and thus determined the means of the data processing by embedding such Like-button to its website.[74] As the GDPR was not applied to this particular case, it is not clear whether the interpretation

---

[68] Article 26(2) of the GDPR.

[69] As criticized in Zalnieriute and Churches (2020) p. 862.

[70] CJEU judgement in case *Fashion ID*, paragraphs 27, 75 and 83.

[71] Opinion of Advocate General Bobek: Fashion ID GmbH & Co.KG v Verbraucherzentrale NRW eV ECLI:EU:C:2018:1039 [2018], paragraph 18.

[72] CJEU judgement in case *Fashion ID*, paragraph 82.

[73] *Ibid*. paragraph 76.

[74] *Ibid*, paragraphs 77–78.

applies to Article 26 of the GDPR as well.[75] Nevertheless, what is noteworthy is that determining the purposes and means of the data processing is not always clear, and unclear situations may lead to unwanted outcomes from the point of view of the (joint) controllers.

3.3 Information obligations of the data recipient when collecting data from other sources

After assessing the information obligation of the data collector as an independent controller, a processor and a joint controller, we move to assess the obligations of the data recipient, to which the data collector aims to sell and transfer the collected (personal) data.

Principle of transparency and principle of purpose limitation limit the processing of data recipient as well. The informational obligations of the recipient of personal data are almost identical than those of the data collector, which has been assessed earlier, as long as the recipient is an independent controller.[76] Therefore, mainly the differences to those obligations are assessed in this chapter. The informational obligations enforce the principle of transparency. Where the data recipient is a processor or a joint controller, the obligations and responsibilities which have been assessed earlier apply.

According to Article 12, where the personal data has been collected from sources other than directly from the data subject, the information of the means of collection as well as data subject's rights should be given to the data subject in reasonable period of time from the collection of the personal data. Where the origin of the personal data is not possible to disclose due to the multiple sources of information, general information should be provided of the source of the personal data.[77]

The information to be provided whether the source of information is the data subject itself (Article 13) or other sources (Article 14) does not differ a lot. Firstly, under Article 14, unlike in Article 13, the controller is not obliged to mention whether providing the personal data is based on legislation or a contract, or whether the data subject is obliged to provide his or her

---

[75] See more of the discussion in Zalnieriute and Churches 2020 p. 869–875.

[76] The similarity of the obligations is visible even from the layout and wording of Articles 13 and 14 of the GDPR. At this point I also want to remind of the vocabulary differences between the GDPR and my paper: whilst the GDPR does not define the entities taking part in the data collection and further disclosing, I use terms data collector and data recipient.

[77] Recital 61 of the GDPR.

personal data, and what are the possible consequences if he or she fails to provide such personal data.[78] The differences with Article 13 are understandable, since the controller in this case has already received the data from a source other than the data subject.

Secondly, under Article 14(2)(f), controllers are obliged to disclose the source of the personal data, and whether the personal data came from publicly available sources. Such information is not needed where the controller has obtained the data directly from the data subject, as is the case in situations referred to in Article 13.

As said, the information under Article 14 should be provided to data subject within a reasonable time, at latest within one month from obtaining the personal data, and where the data is used to communication with the data subject, at the latest when contacting the data subject for the first time, and where the data is disclosed to other recipients, at the latest when it is first disclosed.[79]

The obligation to provide the data subject relevant information in accordance with Article 14 of the GDPR does not apply where the data subject has already received such information.[80] This would happen in situations where the data collector has informed the data subject of all relevant aspects of the data processing, including but not limited to, every applicable processing purpose, each data controller and processor, and in case of joint controllers their distribution of obligations and processing activities, and information on how to exercise the rights of the data subject. However, the controller has the burden of proof that the data subject has already received the information.[81]

---

[78] See, for example, Korpisaari – Pitkänen – Warma-Lehtinen 2018 p. 197.

[79] Article 14(3) of the GDPR.

[80] Article 14(5)(a) of the GDPR.

[81] Korpisaari – Pitkänen – Warma-Lehtinen 2018 p. 182.

4 Relevance of the legal basis for data collection

4.1 Consent

4.1.1 General definitions and requirements

Processing of the personal data must always be lawful, that is the controller must have a justification for processing, which could be any one of the six legal bases that are mentioned in Article 6(1) of the GDPR. Due to limits of this paper, I exclude those legal bases that are mostly available to public bodies, and assess only three out of six legal bases: consent, performance of a contract, and pursuing legitimate interests of the controller or of a third party, and only in an essence that is relevant from the point of view of the research topic of this paper.[82]

Burden of proof of the existence and legitimacy of the justification is of controllers. Controller may rely on several bases simultaneously for different groups of personal data or different purposes of processing, and bases do not have a presumed order of priority.[83] However, the basis of the processing should be decided beforehand, since the controller cannot swap from consent to other bases, where, for example, there has been problems with the validity of consent.[84]

First and foremost, as practically every EU-citizen has noticed, the most relied justification for processing is consent, which means, under the definition provided in the Article 4(11), "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". According to Article 6(1)(a) data processing is lawful when the data subject has given consent to the processing of his or her personal data for one or more specific purposes.

---

82 Other three legal bases for processing personal data are compliance with a legal obligation, vital interests as well as public interest. Nonetheless, those legal bases should not be used regularly by private and commercial entities. See for example Voigt – von dem Bussche 2017 pp. 92–115 for more comprehensive analysis on those legal bases for data processing.

83 Korpisaari – Pitkänen – Warma-Lehtinen 2018 p. 100.

84 WP259 rev.01 p. 23. For contradictory opinion see Voigt – von dem Bussche 2017 p. 101, where authors state that the practice in which several bases has been used in collecting personal data can be upheld under the GDPR.

4.1.2 Obtaining and revoking consent

In practice the most trustworthy and relied legal basis for processing personal data is consent of the data subject[85], usually by obtaining it through pop-up banners at the time when the internet user is opening a webpage. Although a whole dissertation could be written on the definition, delimitation and problems of consent, I limit my research to answer the following questions: a) could consent be effectively obtained from the data subject if the personal data are collected to be disclosed to other recipients either directly or later, b) can the data recipient rely on data subject's consent which has been obtained previously by the original data collector, and c) how can one revoke consent and should it be informed to both controllers?

Firstly, obtaining consent from the data subject can take any form, but consent must always be given freely, informed and separately for each processing purpose, including disclosing the personal data to other recipients, if the processing activity relies on justification of data subject's consent.[86] What is more, controllers cannot bundle data subject's consent to cover all processing activities and purposes, but instead every processing purpose must be specified, and there must be a possibility to separately consent to each of the purposes.[87]

Thus, if one of the purposes is to collect and disclose personal data to specified and identified or identifiable recipients, the consent could be obtained effectively, presupposing that other conditions for consent are met.

Secondly, when it comes to the coverage of such consent that has been given to the *data collector*, it must be noted that the consent only covers the data processing activities and purposes of that specific collector, that is, the *data collector*.[88] Consequently, the *data recipient*, to whom the personal data are being disclosed, must be able to demonstrate that they possess a legal basis for data processing as well, be it consent, contract, or legitimate

[85] Edwards 2016 p. 53; Kramcsák 2023 p. 5.

[86] Recital 32; WP259 rev.01 p. 9-10.

[87] Recital 32. See also European Data Protection Board (EDPB) Guidelines 05/2020 p. 12.

[88] According to Article 7(1) of the GDPR "the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data." This should be interpreted that the controller cannot permit other entities to rely on the consent obtained from the data subject as well, and thus all data recipients, which are independent controllers, must obtain the consent as well (or rely on other legal justifications, such as pursuing legitimate interests of the controller *or third parties*).

interests (the last two will be discussed later in this paper). Therefore, the *data recipient* must also possess a legal basis for the processing to being able to process the personal data.

However, the European Data Protection Board (EDPB) has stated regarding the conditions of consent that "where consent sought is to be relied upon by multiple (joint) controllers or if the data is to be transferred to or processed by other controllers who wish to *rely on the original consent*, these organisations should all be named" (italics added)[89]. Even though the EDPB continues by stating that "more information may be needed to allow the data subject to genuinely understand the processing operations at hand", I cannot but criticize their point of view: I am of a view that each and every controller should obtain and possess their own, separate lawful basis for data processing. I understand this might cause compliance costs, but it is against the GDPR principles, especially the principle of accountability under Article 5(2) of the GDPR, that the controller shall be responsible for, and be able to demonstrate compliance with the GDPR obligations. Usually, if the processing purposes are legitimate and lawful, those data recipients could rely upon the legal basis of the legitimate interests, which would be more legitimate than to rely upon consent of the data subject, who has given the consent to *another* controller.

Even if the controllers to whom the data are disclosed and who are relying upon the original consent of the data subject, they must be able to demonstrate the existence and validity of the consent of the data subject.[90]

Processors (who are processing personal data *for* the controller) can rely on the consent obtained by the controller. To fulfil the information obligations as discussed earlier, those processors should also be identified and informed the data subject of.[91]

Either way, transferring personal data to other recipients requires a separate consent from the data subject (and under information obligation to disclose the recipient(s)), be the recipient a processor, a joint controller or an independent controller.[92] Where other (joint or independent)

---

[89] European Data Protection Board, Guidelines 05/2020 on consent under Regulation 2016/679, paragraph 65 p. 16.

[90] EDPB Guidelines 05/2020 paragraph 104 p. 22; Recital 42 of the GDPR.

[91] EDPB Guidelines 05/2020 paragraph 65 p. 16.

[92] EDBP Guidelines 05/2020 paragraphs 42–45, p. 12.

controllers are relying on the original consent, those entities should all be named when obtaining the consent.[93]

Thirdly, the data subject must have the right to withdraw his or her consent at any time.[94] Revoking the consent must be as easy as it was to give it, although not necessarily exactly the same way.[95] Where the consent has covered several processing purposes and several data recipients (processors, joint controllers, or independent controllers), the wording of the Article would suggest that it should be enough to inform the original controller (in here, the *data collector*) of the revoking of the consent, and such withdrawal would cover the processing activities of the data recipients who have been processing the personal data of the data subject relying upon his or her consent, even though the EDPB guidelines and the GDPR do not provide more specific information on this topic.

After the data subject has revoked the consent, those processing purposes which relied upon the consent of the data subject shall be halted and the collected data irreversibly anonymized or deleted.[96] Where the processing has had other legal basis as well to process such data, there is no obligation to delete such data.[97]

4.1.3 Detriment

Withdrawing or refusing consent from one or several processing purposes must no lead to detriment of the services provided or other costs to the data subject.[98] In other words, the service must remain the same or almost the same even after refusing to give the consent, or after the withdrawal of the consent. Not all incentives to obtain consent are excluded, but the controller must demonstrate that consent was given voluntarily.[99]

---

[93] *Ibid*.

[94] Article 7(3) of the GDPR.

[95] EDPB Guidelines 05/2020 paragraphs 113–114, p. 23–24.

[96] EDPB Guidelines 05/2020 paragraph 117 and 119, p. 24–25.

[97] To get clearer picture of the functioning of the withdrawal of the consent with other legal bases, see EDPB Guidelines 05/2020 paragraphs 112–123, p. 23–25 as a whole.

[98] Recital 42 of the GDPR.

[99] WP259 rev.01 p. 11.

Several mobile applications and webpages use data analytics tools to measure browsing behavior and consuming habits of webpage visitors or users. Where data analytics tools collect personal data, there must be a legal justification for it, usually consent of data subject. Thus, should data subject refuse to give his or her consent to data analytics, where the personal data are shared with other (disclosed) recipients, the refusal shall not affect the user experience in any significant negative manner, such as higher prices.

4.1.4 Function creep and consent fatigue

Consent may cover different processing activities as long as those activities serve the same purpose for which the consent was originally obtained to. If the personal data that have been collected are being processed for other purposes than of those of which the data subject has been informed of, it may result in breach of Article 6 and significant administrative fines to the controller. The phenomenon, in which the data are collected extensively and gradually and furthermore being processed to other purposes than the data subject had been informed of, is called function creep.[100]

As many may have noticed, consenting to data collection several times a day may lead to consent fatigue, where people no longer care about protecting their personal data, and just consent to almost everything.[101] GDPR problematically has not tackled the issue properly and relies on controllers to solve problems. Controllers have not incentive to solve the issue, since they can rely on data subjects' consent quite easily, when they have a proof in writing or in electronic form that those data subjects have consented to their data collection, even though only a few has actually read through the terms of data collection properly.

In principle, each controller must be able to rely on its own justification for processing of personal data. Where the consent has been obtained from the data subject for the personal data to be transferred to other recipients, has at least the data collector fulfilled its obligation regarding the legal justification for processing. On the other hand, or on the other side of the data transaction, to proof the existence of a valid consent may be more difficult to do for the

---

[100] More of function creep in the GDPR framework, see WP259 rev.01 p. 12, and of the term and its definition, see Koops, Bert-Jaap. "*The Concept of Function Creep*." Law, innovation and technology 13.1 (2021): 29–56, where Koops defines the term as being "an imperceptibly transformative and therewith contestable change in a data-processing system's proper activity."

[101] WP259 rev.01 p. 17.

recipient of data, unless it specifically asks the data subject his or her consent to processing the personal data, or the consent has already been given for transferring the data to mentioned recipients.[102]

Explicit consent

An explicit consent is needed to for processing special categories of personal data as meant in Article 9[103]; to data transfers to third countries[104], as well as to use automated decision-making, including profiling, as meant in Article 22[105]. Data subject must expressively give a statement of consent where needed, although the form is not tied to any specific form.[106]

In sum, when collecting personal data and sharing them with other data recipients, consent is a valid justification so long as the consent is actually freely given, with an actual possibility to opt-out from processing purposes that are not necessary under other legal bases for processing. The recipient of the personal data must be able to demonstrate it also has the consent for processing, whether obtained by itself or by the original controller (here: data collector).[107] Due to limits of this paper the processing of special categories of personal data, to which the explicit consent is tied to, is not specifically addressed, unless it has been relevant for the context.

4.2 Performance of a contract

4.2.1 General conditions

Under Article 6(1)(b) processing is lawful when "procsessing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the

---

[102] WP259 rev.01 p. 10.

[103] Contractual justification does not apply when processing special categories of personal data, hence controllers should rely either on exceptions under Article 9(2) subparagraphs (b) to (j) or an explicit consent.

[104] An explicit consent can justify data transfers to third countries without adequate level of data protection, see Article 49.

[105] Due to research economic reasons automated decision-making is mostly left out of the scope of this paper. However, there are lots of material regarding automated decision-making and collection of data thereof. See "WP29 Guidelines on Automated decision-making and Profiling for the purposes of Regulation 2016/679" (WP 251).

[106] For more information with examples on explicit consent, see WP259 rev.01 pp. 18–20.

[107] EDPB Guidelines 05/2020 paragraph 104 p. 22.

request of the data subject prior entering into a contract". In other words, this legal basis fits well in customer relationships, in which processing of personal data is necessary, for example, to ship the goods to buyer. The extent to which the data are processed should be limited to the extent that is strictly necessary for the contract to be performed, that is if the contract cannot be fulfilled without the processing of such data.[108] For example, when delivering goods to the customer, processing of the address and credit card information would be legitimate under this legal basis.[109]

The legal basis of performance of a contract applies when either of the following two conditions are met: the processing is objectively necessary for the performance of a contract to which the data subject is a party, or the processing is objectively necessary in order to enter into a contract to which the data subject is a party, at the request of the data subject.[110]

The necessity condition is met when the there are no other less-intrusive options than to process the data to fulfil the contract.[111] Other legal bases should be considered where the processing is useful but not objectively necessary for performance of the contract, even when taking place at the request of the data subject, or where the processing is merely mentioned in the contract without being necessary.[112]

4.2.2 Limitations to the use of performance of a contract as a lawful basis

As can be observed, relying on the legal basis of performance of a contract is not without limitations. In addition to the necessity limitation, the purpose limitation principle as well as the data minimization obligation limit the use of this legal basis.

When collecting personal data to create excessive datasets to be disclosed to other recipients, this legal basis does not fit well, since the data subject must be a party to a contract, or prior to

---

[108] Voigt – von dem Bussche 2017 p. 102.

[109] See, for example, EDPB Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects, version 2.0, 8 October 2019, Example 1, p. 10.

[110] EDPB Guidelines 2/2019 paragraph 22, p. 8.

[111] EDPB Guidelines 2/2019 paragraph 25, p. 8; CJEU, Joined Cases C-92/09 and C-93/09, Volker und Markus Schecke GbR and Hartmut Eifert v Land Hessen, 9. November 2010.

[112] EDPB Guidelines 2/2019 paragraphs 25, 27, pp. 8-9; Article 29 Working Party Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC (WP217), page 16–17, endorsed by the EDPB.

entering into a contract to happen at the request of the data subject. Where one of the purposes of the data processing is the personal data are being disclosed to third parties, instead of fulfilling necessary contractual obligations of a contract to which the data subject is a party, another legal basis should be found to it, for example a consent of the data subject or legitimate interests of the controller or of a third party.

EDPB has also held that relying on the contract basis is hardly applicable to social media providers, as the advertisement partly and indirectly funds their services, and thus other legal bases should be considered.[113]

An important part of insurance companies' business is collecting relevant information about their customers. To what extent, then, can they collect personal data on the legal basis of performance of a contract? One could assume that at least to collect name and personal identification or social security number would be justified to perform the insurance contract. How about credit information or address details from outside sources, such as a bank or a register? If those data are necessary for the contract, and the processing takes place at the request of the data subject prior entering into a contract, it should be lawful. On the other hand, collecting behavioral information from third-party sources about the data subject should not be justified under this legal basis.

Also, processing of sensitive personal data as meant in Article 9 of the GDPR is never justified under the legal basis of performance of a contract.[114]

4.3 Legitimate interests

4.3.1 General conditions

In addition to consent and contractual basis, legitimate interests could also be used as a legal basis when collecting personal data of the data subject from third parties. According to the Article 6(1)(f), processing may be lawful when “processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such

[113] EDBP Guidelines 2/2019 paragraphs 52 and 53.

[114] Article 29 Working Party Guidelines on consent under Regulation 2016/679 (WP259), endorsed by the EDPB, page 19.

interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child".

Legitimate interests as a legal basis provides flexibility to the interpretation of the GDPR, since not all processing activities can be predicted at the time of assessing suitable legal bases for processing.[115] Legitimate interests are not strictly defined in the GDPR, but while the legal basis of legitimate interests existed already in the GDPR predecessor Data Protection Directive, the Data Protection Working Party (WP29) defined that a legitimate interest shall "be lawful (i.e. in accordance with applicable EU and national law), be sufficiently clearly articulated to allow the balancing test to be carried out against the interests and fundamental rights of the data subject (i.e. sufficiently specific), [and] represent a real and present interest (i.e. not be speculative)."[116] Moreover, an interest is a broad stake or benefit that the controller derives from the processing.[117]

To use the legal basis of legitimate interests, controllers' must be able to proof 1.) they possess a legitimate interest to be protected; 2.) that such legitimate interest is lawful. clear and real, 3.) the processing of personal data is necessary to protect such interests, and that 4.) such legitimate interests are not overridden by the interests and rights of the data subject.[118] Thus, the controller carries the burden of proof for the existence and legitimacy of legitimate interests that are not overridden by the data subject's interests or rights.

In practice, legitimate interests of the controller may be of variable nature, such as legal, ideological, or even commercial[119]. Recital 47 of the GDPR has specifically referred to direct marketing purposes as an example of legitimate interests as a legal basis for processing of personal data to highlight the fact that the nature of the legitimate interest may vary a lot. What is more, the Data Protection Working Party has directly mentioned "conventional direct marketing and other forms of marketing or advertisement" when giving examples of cases

[115] Korpisaari – Pitkänen – Warma-Lehtinen 2018 p. 117.

[116] WP 217 p. 25.

[117] Ibid. p. 23.

[118] Korpisaari – Pitkänen – Warma-Lehtinen 2018 pp. 119-122; Office of The Data Protection Ombudsman (Finnish Data Protection Authority) at https://tietosuoja.fi/en/controller-s-legitimate-interests, accessed February 16, 2023.

[119] Voigt – von dem Bussche 2017, pp. 269-270.

(although not including the balancing test assessment) where the legal basis of legitimate interests may be used.[120]

Legitimate interests may also be of third parties, as it is explicitly mentioned in the wording of Article 6(1)(f) as well as in Recital 47. There seems not to be any restrictions regarding whose legitimate interests may be protected under this Article, but, however, the principles of the GDPR and the balancing test, which is assessed next, under the legitimate interests basis restrict the use and interpretation of this legal basis.

After the assessment of applicability of the legitimate interests in data processing, the controller is not free of other requirements stemming from the GPDR. The mere fact that some sets of personal data is allowed to be processed does not mean that there are not limits in data processing. The principles of data minimization as well as purpose limitation remain in restricting the processing of such data.

4.3.2 Balancing test

As the wording of the Article 6(1)(f) suggests, the interests or fundamental rights and freedoms of the data subject, which require protection of personal data, restrict the processing of personal data of the data subject under the legal basis of legitimate interests. The assessment whether those interests, rights or freedoms override those of controller or third party is conducted in a balancing test, which should be done and documented by the controller before processing of personal data under this basis.[121]

The rights, freedoms and interests of the data subject may be of an idealistic, economic, social, professional, private or other nature.[122] The relationship between the data subject and the controller, as well as data subject's reasonable expectations should be taken into account. Where the data subject is a child, the need for privacy protection is stronger.

In the balancing test three aspects need to be assessed: 1.) the pursuit of legitimate interests of the controller of third party, 2.) the necessity of the processing of personal data to realize

[120] WP 217 p. 25.

[121] WP 217 p. 43. See more of the balancing test in Kamara, I. and De Hert, P. (2019). Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach. In Seligner, E., Polonetsky, J., and Tene, O., editors, The Cambridge Handbook of Consumer Privacy. Cambridge University Press

[122] Voigt – von dem Bussche 2017 p. 105.

those legitimate interests, and 3.) that there are no overriding interests, rights or freedoms of data subjects.[123] When weighing the interests against each other, the consequences of the processing activities for the data subject's rights and freedoms are assessed against the legitimate interests of the controller.[124] Where the interests of the data subject override the interests of the controller, the processing is not lawful under the legal basis of legitimate interests.

4.3.3 Reasonable expectations

When assessing whether the legitimate interests is an appropriate legal basis for processing the reasonable expectations of the data subject must be taken into account – it must be assessed whether a data subject can expect his or her data to be processed for specific purpose at the time of the collection of the collection of the personal data.[125] If the data subject cannot reasonably expect his or her data to be processed, it is likely that the interests of the data subject override those of the controller.[126]

Voigt compares the legitimate interests basis and contractual basis when it comes to reasonable expectations of the data subject. Where the legal basis of performance of a contract has been used to personal data processing, but where the processing is exceeding the limit of what is necessary for the performance of a contract, in other words cannot be lawful anymore under the Article 6(1)(b) of the GDPR, it is likely that the data subject cannot reasonably expect the data to be processed under the legitimate interests either.[127]

4.3.4 Legitimate interests of third parties

In Data Privacy Directive Article 7(1)(f), the predecessor to the GDPR, the legitimate interests basis recognized also the legitimate interests of a third party "to whom the data are disclosed". It is noteworthy that the text has been modified in the GDPR, and it does not refer only to parties that the data are (or have already been) disclosed to. Initially, the GDPR was

---

[123] Voigt – von dem Bussche 2017 pp. 105–106; and with reference to the equivalent provision in Article 7(f) of Directive 95/46, the judgment of 4 May 2017, Rīgas satiksme (Case C-13/16, EU:C:2017:336) paragraph 28.
[124] Voigt – von dem Bussche 2017 p. 106.

[125] Recital 47 of the GDPR.

[126] Ibid.

[127] Voigt – von dem Bussche 2017 p. 106.

not meant to even mention third parties, but however they were reintroduced to the provision.[128]

Thus, technically there are no fundamental limits when assessing the legitimate interests of third parties when collecting personal data from other sources than the data subject him- or herself. Therefore, in some cases even technology giants could rely on their legitimate interests (such as marketing purposes, more efficient directed marketing, their own commercial benefits) when collecting the personal data of their customers from data collectors, other online services and webpages, and even "offline" companies and organizations, subject to passing the balancing test and safeguarding the interests and rights of data subjects.

However, creating such large datasets comprising of large amounts of personal data may be not reasonably expected by the data subject, and therefore the fundamental rights and freedoms or interests of the data subject may override controller's or third parties' in such cases, especially where those data were initially collected from other sources and for different purposes.[129] What is more, the WP29 has stated that profiling and tracking for advertising purposes would be difficult to justify under the legal basis of legitimate interests.[130] In those situations, consent of the data subject would be more acceptable legal basis.[131]

Legitimate interests as a legal basis for personal data processing cannot be used when collecting user data with cookies, but instead a user consent is required, according to the Article 5(3) and Article 13 of the ePrivacy Directive, as well as the case law of the CJEU[132]. Since the ePrivacy Directive falls out of the scope of this paper, I am not assessing it further here.[133]

---

[128] WP 217 p. 27. See also footnote 54 at the same page.

[129] WP 217 p. 25.

[130] Article 29 Working Party, Opinion on profiling and automated decision-making, WP 251 rev. 01 p. 15; WP 217 p. 32.

[131] EDPB Guidelines 08/2020 on the targeting of social media users, 7 July 2021, paragraphs 56–58 p. 18.

[132] Court of Justice of the European Union, Judgment in Planet 49 Gmbh, Case C-673/17, paragraph 73; Fashion ID, C-40/17, paragraphs 89–91.

[133] See more of the interplay between the GDPR and the ePrivacy Directive in EDPB Opinion 5/2019 on the interplay between the ePrivacy Directive and the GDPR.

4.3.5 Additional safeguards

Where it is not clear whether the controller may rely on legitimate interests after it has been conducted the balancing test, it may set up additional safeguards to protect the fundamental rights and freedoms as well as interests of the data subject, and thus ensure the lawfulness of the processing. Additional safeguards could consist of, for example, limitations to the amount and nature of the data being collected and strict limitations to the period of time for how long the data are stored.[134] It should be noted, however, that the EDPB has not updated their guidelines, and the opinion of the Data Protection Working Party may be outdated and not strictly applicable to the legitimate interests basis under Article 6(1)(f) of the GDPR when it comes to additional safeguards.

4.3.6 Data collection to train artificial intelligence

Maybe the most discussed innovation in 2022-2023 has this far been generative artificial intelligence (AI) models. AI, previously being only a buzzword in the language of consultants in recent years, took off and became a hot topic in everyday discussions in the late November 2022 when U.S. company OpenAI published their chatbot called ChatGPT for public use, which is a general-purpose chatbot powered by large language model (LLM) called GPT-4.[135]

Not only did the public get interested in the chatbot and increased development of AI, but the data privacy experts did so as well. While AI poses some risks to individuals, maybe even societies, those could and should be tackled. From the point of view of data privacy and this paper, it is of utmost importance to scrutinize into the data collection phase of AI development and the legal basis for data processing.

Development of AI innovations, such as ChatGPT, require access to large and veracious datasets. Those large and labelled datasets are critical to the success of machine learning (a part of AI), including features of timeliness and representativeness.[136] On the other hand, if

---

[134] WP 217 pp. 42–43.

[135] For more information of the ChatGPT and its recent developments, see an updating article of TechCrunch: ChatGPT: Everything you need to know about the AI-powered chatbot by Alyssa Stringer and Kyle Wiggers, https://techcrunch.com/2023/04/11/chatgpt-everything-you-need-to-know-about-the-ai-powered-chatbot/ (accessed April 13, 2023).

[136] Pablo Trigo Kramcsák, "Consent and Legitimate Interests as Legal Bases for Data Processing Under the GDPR", Computer Law & Security Review 48 (2023), p. 4.

the data that has been used to train the algorithm is of bad quality, it affects the functioning of the algorithms and thus may lead to untruthful, wrong, implausible or even dangerous outcomes.[137]

Thus, in order to develop functional AI models, algorithms need training data. Datasets which are of an adequate quality often include personal data, and as we know, the GDPR applies where the data processing or the data subjects of such processing are located in the EU/EEA. The obligations and possible restrictions of the GDPR might be of a problem for AI developers and future innovation, at least from the commercial point of view, as the AI is in many cases reliant on third-party data: creation of Large Language Models (LLM) would be extremely slow if the developer organization should create the datasets themselves. Scraping data from internet sources is more effective.

Nonetheless, the rights and freedoms of individuals are in the core of data protection law, that is the right to privacy and right to data protection[138], but there are also other foreseeable and unforeseeable problems arising from rapidly growing use of AI, such as mistakes of AI caused by false or wrongfully adapted training data or other sources the AI is using.

Leaving the problematization of the AI behind for a while (I will assess the phenomenon in more depth in the last section of this paper), let us focus on the legal basis of data collection in order to develop AI systems for a while. AI developers may collect training data from various sources and rely on one of the following legal bases: consent of a data subject, performance of a contract, or legitimate interests of a controller.

Given the characteristics of Big Data[139], consent of the data subject, even though being "the most global standard of legitimacy… and most likely to engender user trust"[140], when training AI algorithms with large datasets (Big Data) could turn very difficult, since 1.) there could be

---

[137] Some describe the phenomenon aptly as a shit in/shit out effect.

[138] See European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, COM (2020) 65 final, p. 11, available at https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en, accessed 27 April 2023.

[139] Like AI, Big Data lacks a clear legal definition as well, but could be defined with three V's: Volume, Velocity, Variety. Big Data is often used in training AI models. See for example S. Sagiroglu and D. Sinanc, "Big data: A review," 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 2013, pp. 42-47.

[140] Lilian Edwards, 'Privacy, security and data protection in smart cities: a critical EU law perspective' (2016) 2(1) European Data Protection Law Review 28, p. 53.

an unlimited number of data subjects, 2.) obtaining a clear and voluntary consent for data processing would be challenging, if not impossible, and what is more, 3.) if consent is revoked, it may be impossible to stop processing the personal data which previously was allowed, and to "unlearn" the algorithm and take off the data from the servers[141].

Moreover, 4.) keeping in mind that when obtaining the consent, the purposes of the data processing must be defined, and the consent for each of those purposes must be obtained separately[142]. Thus, it is not always possible to foresee or understand at the time of obtaining the consent, that the personal data would be used to develop AI models.[143]

As we can see, consent of a data subject does not fit to the AI development very well. The legal basis of performance of a contract is at least as problematic. In the spring of 2023, the chatbot ChatGPT of OpenAI got its first (and presumably not last) drawbacks in the EU for not complying with the GDPR, at least according to the Italian Data Protection Authority *Garante per la protezione dei dati personali* (hereinafter Italian SA), which held in its decision that in order to comply with the GDPR, OpenAI cannot use the legal basis of performance of a contract (in this case the terms of use of the service), and thus must rely on either consent of data subjects of legitimate interests of the controller or of a third party as their applicable legal basis for data processing of their Italian users.[144] Until the ChatGPT complies with the GDPR, the Italian SA will keep in force their temporary limitation on the processing of Italian users' personal data, affecting the service to temporarily be shut down in Italy.[145]

---

[141] Kramcsák 2023 p. 5.

[142] *Ibid* p. 5-6. According to the Article 6(4) of the GDPR, if the requirements of the reuse of the personal data (the processing must be compatible with the purposes for which the consent was originally obtained to) are not met, a new consent would be required.

[143] *Ibid*.

[144] ChatGPT: Garante privacy, limitazione provvisoria sospesa se OpenAI adotterà le misure richieste, L'Autorità ha dato tempo alla società fino al 30 aprile per mettersi in regola (in English: ChatGPT: Italian SA to lift temporary limitation if OpenAI implements measures, 30 April set as deadline for compliance). A press release by the Italian SA on 12 April 2023. Available at https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9874751 , accessed on 14 April 2023.

[145] Intelligenza artificiale: il Garante blocca ChatGPT. Raccolta illecita di dati personali. Assenza di sistemi per la verifica dell'età dei minori (in English: Artificial intelligence: stop to ChatGPT by the Italian SA, Personal data is collected unlawfully, no age verification system is in place for children) A press release by the Italian Data Protection Authority on 31 March 2023. Available at

Moreover, keeping in mind what has already been said about the hardships of the performance of a contract as a legal basis for data processing, using it to develop AI models could be used only in very narrow situations. Therefore, legitimate interests of the controller would be more useful in AI development.[146]

AI development phase, including supplying the algorithm new training data in a closed environment (supervised training)[147], should be separated from the actual use (application) of those AI models[148]. Development phases bear a different risk position in comparison to the application phase.

According to Hacker, a more permissive interpretation of the legal basis of legitimate interests would be welcoming in order to enable AI model training. In his view, training the datasets does not reveal anything substantially new about the default risk of the data subjects in relation to those datasets.[149] However this does not come without exceptions: if the data is of sensitive nature (as meant in the Article 9 of the GDPR), or if there is a need to transfer or disclose the data to new controllers during the training phase, a more restrictive interpretation should be applied in the balancing test of the legitimate interests.[150]

In the training phase of the AI model, legitimate interests of the controller and of third parties are weighed against rights and freedoms of those data subjects who are represented in the training data sets. Key factors to mitigate the risks and to tip the balance to the benefit of the AI developer (data controller) are anonymization[151], societal benefits of the AI development, limiting the data storage periods, and avoiding the use of sensitive personal data.[152]

---

https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english, accessed 14 April 2023.

[146] Philipp Hacker, 'A Legal Framework for AI Training Data' (2021) 13(2) Law, Innovation and Technology 257, p. 291.

[147] For a brief overview of the AI training phases, see Hacker 2021 p. 258-259 with comprehensive sources.

[148] Hacker 2021 p. 291.

[149] Hacker 2021 p. 291–292.

[150] Hacker 2021 p. 293.

[151] Hintze, 'Viewing the GDPR Through a De-identification Lens: A Tool for Compliance, Clarification, and Consistency' (2018) 8 International Data Privacy Law 86 p. 94.

[152] Article 29 Working Party, 'Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC', 2014, WP 217, p. 37.

As we have learnt, the legitimate interests could be of any nature, and not only benefit the commercial interests of the AI developer (controller), but in the best-case scenario also the society as a whole[153], and therefore the legal basis of legitimate interests suits the AI development, also when the data is collected from other sources than the data subjects themselves, which is usually the case.

It even has been examined that training AI models (presumably using supervised training practices) would not affect the interests of data subjects, as long as strong security measures have been applied as well, that is usually pseudonymization and anonymization of the personal data, even when those legitimate interests of the controller would be commercial.[154]

Given the difficulties of the anonymization and pseudonymization practices this statement seems a little too bold – if re-identification is possible even after pseudonymization or anonymization[155], those issues should be taken into account in the balancing test, instead of giving a green card to all AI development practices under the umbrella of legitimate interests.

---

[153] However, in the balancing test only the interests of the controller or of a third party and of the data subject are taken into account, not general interests. Those general interests could, however, tip the balance to the benefit of the controller.

[154] 'The impact of the general data protection regulation on artificial intelligence', European Parliament, Directorate General for Parliamentary Research Services, 2020, p. 50.

[155] See, for example, Sweeney, 'Uniqueness of Simple Demographics in the U.S. Population, Laboratory for International Data Privacy' (2000) Working Paper LIDAP-WP4; and Narayanan and Shmatikov, 'Robust De-anonymization of Large Datasets' (2008) Proceedings of the 2008 IEEE Symposium on Security and Privacy 111.

5 The German Facebook Case

5.1 Introduction, the legal issues at stake

As referred to earlier, the so-called German Facebook case illustrates both the position of the data protection law in between the competition law and the consumer law, but also the hardships in ensuring compliance with the GDPR when collecting data from third parties.

However, as this is not comprehensive research on the compliance with the GDPR, nor a competition law study, no conclusions will be drawn regarding the general GDPR compliance of the case or the competition law related aspects. Nevertheless, a short case law analysis regarding the business practices of data collection under the GDPR is needed in here.

Evaluation of the case has sought to highlight the limitations of GDPR in relation to the collection of personal data from parties other than data subjects, focusing on the challenges of three different legal bases in a practical context.

The German Facebook case[156], a case between Meta Platforms, Inc. (formerly known as Facebook, Inc., hereinafter "Meta") and Bundeskartellamnt, German Federal Cartel Office, hereinafter "FCO". The case started with FCO's decision by which the FCO prohibited Meta from processing data and implementing their terms of service in Germany.[157] Meta appealed to the Higher Regional Court of Düsseldorf, which then referred the case to the Court of Justice of European Union (hereinafter "CJEU") to ask for a preliminary ruling. In the CJEU the case is still pending, but the Opinion of Advocate General Rantos is published on 20 September 2022[158], and the preliminary ruling of the CJEU is expected to be delivered in 2023.

In its decision the FCO prohibited Meta implementing their terms of service, since it found them to be contradictory to German competition law, namely as an abusive exploitation of a

[156] Kerber, Wolfgang, and Zolna, Karsten K.. 'The German Facebook Case: The Law and Economics of the Relationship between Competition and Data Protection Law'. European Journal of Law and Economics 54, no. 2 (October 2022): 217–50.

[157] Decision B6-22/16 of the FCO on 6 February 2019.

[158] Opinion of Advocate General Rantos delivered on 20 September 2022 concerning the case C-252/21, Meta Platforms Inc v. Bundeskartellamt, ECLI:EU:C:2022:704).

dominant position under the Paragraph 19(1) of the Gesetz gegen Wettbewerbsbeschränkungen (Law against restrictions on competition; 'the GWB')[159], and as a proof of that, a reference was made to the GDPR, and found that the business practices of Facebook were not in compliance with the GDPR. The referring court, that is the Higher Regional Court of Düsseldorf, however, had doubts whether a national competition authority, such as the FCO, may monitor the compliance of the GDPR with an order to end the infringements, so they asked the CJEU for help.[160]

The case is groundbreaking in its attempt to combine the data protection law and the competition law, but is, undeniably, blurring the lines between the data protection and the competition law. As the preliminary decision of the CJEU is still pending, some weight could be put to the AG Opinion of the case: Advocate General Rantos finds the questions regarding the admissibility of competition law authority using the GDPR as a proof of uncompetitive practice being irrelevant, since the national competition authority was not exceeding its competence because it did not try to use the competences as laid down in the GDPR.[161]

5.2 Legal basis for data collection

5.2.1 Questions referred to the CJEU

One of the key issues in the proceedings of the German Facebook case, despite the fact that it was a competition law case in the first place, was the compliance of the data collection practices with the GDPR. Meta had relied on the legal bases of performance of a contract, their legitimate interests, as well as consent. The Higher Regional Court of Düsseldorf questioned their business practices and asked the CJEU their preliminary ruling (among other questions) of the following questions (italics added):

> 3) *Can an undertaking*, such as Facebook Ireland, which operates a digital social network funded by advertising and offers personalised content and advertising, network security, product improvement and continuous, seamless use of all of its group products in its terms of service, *justify collecting data for these purposes* from other group services *and third-party websites and apps* via

[159] See, for example the AG Rantos' Opinion 2022, paragraph 11, p. 4.

[160] Question (1)(a) of the Higher Regional Court of Düsseldorf to the CJEU, as referred to in the AG Rantos' Opinion, paragraph 13, p. 4.

[161] The questions regarding the competence of authority which were referred to the CJEU were questions 1 and 7.

integrated interfaces such as Facebook Business Tools, or via cookies or similar storage technologies placed on the internet user's computer or mobile device, *linking those data with the user's [Facebook] account and using them, on the ground of necessity for the performance of the contract under Article 6(1)(b) of the GDPR or on the ground of the pursuit of legitimate interests under Article 6(1)(f) of the GDPR*?

(4) In those circumstances, can

- the fact of users being underage, vis-à-vis the personalisation of content and advertising, product improvement, network security and non-marketing communications with the user;

- the provision of measurements, analytics and other business services to enable advertisers, developers and other partners to evaluate and improve their services;

- the provision of marketing communications to the user to enable the undertaking to improve its products and engage in direct marketing;

- research and innovation [in the public interest], to further the state of the art or the academic understanding of important social issues and to affect society and the world in a positive way;

- the sharing of information with law enforcement agencies and responding to legal requests in order to prevent, detect and prosecute criminal offences, unlawful use, breaches of the terms of service and policies and other harmful behaviour;

also constitute legitimate interests within the meaning of Article 6(1)(f) of the GDPR if, for those purposes, the undertaking links data from other group services and from third-party websites and apps with the user's [Facebook] account via integrated interfaces such as Facebook Business Tools or via cookies or similar storage technologies placed on the internet user's computer or mobile device and uses those data?[162]

In other words, the appealing court asked the CJEU whether Meta could legitimately rely on the legal justifications of a) performance of a contract and b) pursue of legitimate interests. Moreover, the fifth question which was referred to the court was about other legal bases under Article 6(1) subparagraphs (c)-(e)), and the sixth question whether consent could have been

162 Request for a preliminary ruling by the Higher Regional Court of Düsseldorf regarding the case C-252/21, 24 March 2021, pp. 2–5.

given effectively, as meant in Article 6(1)(a) and Article 9(2)(a) of the GDPR, and freely, as meant in the definition of consent in the Article 4(11) of the GDPR, in those circumstances.

5.2.2 Performance of a contract

The third question referred to the CJEU was whether an undertaking may justify data collection from other intra-group services as well as third-party websites and apps and consequently link the data with user accounts (profiles) on the ground of necessity for the performance of a contract or legitimate interests. Here we are digging deeper into the first part, contractual justification.

In their Opinion, the Advocate General implicitly criticizes the referring court for their ambiguous and vague wording in the question three, and starts with stating that without a detailed case-by-case analysis of the Meta's terms of service "it is impossible to establish whether - - an undertaking - - can comprehensively rely on all (or some) of the grounds set out in the Article 6(1) of the GDPR".[163] According to the FCO the controller shall establish which data is processed under which legal basis, while Meta states that FCO cannot rule that the practice at issue[164] *might* be based on those grounds and therefore it could not be concluded that the practice at issue would be infringing the requirements of the GDPR.[165]

According to the analysis of the Advocate General, the processing has been carried out on the basis of the general conditions of the contract (Meta's terms of service), without the consent of the data subject. That is why, in his opinion, a strict interpretation of the legal basis is needed.[166]

Even though the controller is responsible for demonstrating the compliance with the GDPR, including specifying the purposes and the legal basis of the processing, Meta had not met those requirements.

---

[163] AG Rantos' Opinion 2022, paragraph 50, p. 10.

[164] By "the practice at issue" AG Rantos means the practice of collecting data from other intra-group Meta services as well as third parties, and combining the data with user profiles. AG Opinion paragraph 10, p. 4.

[165] Footnote 70 of the AG Opinion.

[166] AG Opinion, paragraph 51, p. 10. This was also referred to in the request for a preliminary ruling of 24.3.2021 regarding case C-252/21, p. 7, by the Higher Regional Court of Düsseldorf.

As already mentioned earlier, the processing of the personal data under the legal basis of performance of a contract is lawful only and insofar as it is necessary for the contract to which the data subject is party, or prior entering into a contract to which the data subject is a party at the request of the data subject. According to the case law of CJEU, processing under the legal basis of contract is not automatically *necessary* if the processing is useful (instead of necessary) for the activities related to the contract[167], or only carried out at the time of performance of a contract. Rather, it is lawful when it is objectively necessary and there are no other alternatives to fulfil the contractual obligations than to process the personal data. Such necessity could occur for example when an online store delivers their products directly to their customers, and in order to deliver their products, they must process the address and name of the customer.[168]

Advocate General Rantos found it difficult to understand why the personal data that has been processed would be necessary for the performance of the contract, the contract being the terms of use of the service,[169] nor to take steps at the request of the data subject to enter into a contract[170]. Most critical note that Advocate General gave was about the fact that the personal data was collected outside the platform, from third parties. Thus, a specific consent would be required. Nor did the Advocate General find the personal data collected from other intra-group services necessary for the performance of the contract.[171]

As a conclusion in reference to using the legal basis of performance of a contract, according to the Advocate General the processing shall be "objectively necessary for the provision of services relating to the Facebook account"[172], which he found hard to demonstrate. The opinion supports the point of view of the EDPB.[173]

---

[167] EDPB Guidelines 2/2019, paragraph 25.

[168] See, for example, Korpisaari – Pitkänen – Warma-Lehtinen 2018 p. 102.

[169] AG Opinion paragraph 56, p. 11.

[170] AG Opinion Footnote 73, p. 23.

[171] Ibid., para 57, p. 11.

[172] Ibid., paragraphs 70 and 78.

[173] EDPB Guidelines 2/2019 paragraphs 52 and 53, EDPB Guidelines 08/2020 paragraph 49.

5.2.3 Legitimate interests

5.2.3.1 Balancing test regarding the legitimate interests justification

In addition to consent and performance of a contract, in the German Facebook case the third legal basis of which legitimacy was assessed was the legitimate interests of the data controller or third parties under the Article 6(1)(f) of the GDPR.

The Higher Regional Court of Düsseldorf had mentioned three legitimate interests of Meta regarding the processing of personal data, being personalization of advertising, network security and product improvement.[174] At least the first interest could be interpreted directly being a pure economic interest of Meta, the other being more security-related interest, and the last, depending on the actual interests of Meta and the perspective, security or economic-related, or both.

However, Advocate General Rantos was critical towards the legitimate interests justification in this case. Regarding the first interest, personalization of advertising, he found it hard to demonstrate the necessity of the processing, since the origin of the data was third-party apps and websites, and therefore the impact of the processing on the user as well as his or her reasonable expectations and safeguarding measures used by the controller should be assessed more critically.[175] Without further explanation a reference was made to the Data Protection Working Party Opinion 06/2014 on the notion of legitimate interests (WP 217), where the Working Party states that especially the following key factors should be taken into account when applying the balancing test: a) assessing the controller's legitimate interest, b) impact on the data subjects, c) provisional balance and d) additional safeguards. Those aspects were mostly assessed previously in this paper. Here I will assess each factor individually and interpret them to the German Facebook Case.

---

[174] AG Opinion, paragraph 63.

[175] Ibid. paragraph 64.

5.2.3.2 Legitimate interests of Facebook

Firstly, the characteristics of the legitimate interest of the controller may be of fundamental rights, such as freedom of expression and information[176], the freedom to conduct a business[177] and the right to property[178], as protected and meant in the Charter of Fundamental Rights in the European Union (CFR) and in the European Convention on Human Rights (ECHR).

In the case at hand, personalization of advertising, or rather economic or business interest, could fall into under the freedom to conduct a business. As fundamental rights are not unlimited or without any restrictions, the data processing activities conducted by the controller to pursue its legitimate interests must be necessary and proportionate to exercise its fundamental right.[179] In other words, as mentioned earlier, there must not be less intrusive options reasonably available to pursue such legitimate interests other than the processing activities at question, that is, in this case, to conduct personalization of the advertising, and moreover, the processing must be in balance with the rights and freedoms of the data subject.

On the other side of the balancing test the rights and freedoms of the data subject are taken into account. Most evident of those are right to private life[180] and right to protection of personal data[181], which, depending on the case, may be partly parallel or complementary[182]. As this paper is not assessing the data protection law from the point of view of fundamental rights, I am not going to assess those aspects further than it is reasonable in understanding the balancing test, to which I return at the end of this chapter.

5.2.3.3 Impact on data subjects

Secondly, the impact of data processing on data subjects is a crucial factor in the balancing test, where the nature of the personal data that is being processed, the way it is processed, the

---

[176] Article 11 of the CFR; Articles 9 and 10 of the ECHR.

[177] Article 16 of the CFR.

[178] Article 17 of the CFR; Article 1 of the Protocol 1 to the ECHR.

[179] WP 217 p. 34.

[180] Article 7 of the CFR; Article 8 of the ECHR.

[181] Article 8 of the CFR.

[182] Korpisaari – Pitkänen – Warma-Lehtinen 2018 pp. 13–15.

reasonable expectations of the data subjects as well as the relationship between the data controller and the data subject are relevant factors in the balancing test.[183]

In the impact assessment, which must be distinguished from the data processing impact assessment (DPIA) as meant in the Article 35 of the GDPR, both the negative and positive, as well as potential and real, already materialized or materializing impacts should be considered, where the severity of the consequences and its likeliness contributes to the impact assessment.[184] At the case at hands, those impacts could be, among others, the result of the data being disclosed publicly or to third parties, and the distress of the data subjects resulting from realizing that their personal data have been disclosed, misused or leaked.

The nature of the personal data being processed is one of the most important factors in the balancing test. In the case at hands, there is not (yet, at least) specific information on the nature of the data that was being processed, other than the fact that the Higher Regional Court of Düsseldorf referred to sensitive data in its second question about the legitimacy of the consent, and disclosed that some information were collected from flirting apps, gay dating sites, political party websites and health-related websites.[185] However, these issues were brough up in reference to the processing of special categories of personal data under the Article 9(1) of the GDPR, which is prohibited unless specifically justified under one of the justifications in Article 9(2). Therefore, it cannot be assessed whether sensitive data were collected under the legitimate interests justification. Nevertheless, Article 9(2) of the GDPR does not recognize legitimate interests as one of the legal basis when processing special categories of personal data.

Another aspect of the impact on data subjects is the way data are being processed: is it made available to the public (such as publishing the points of “interests” of the data subject in his or her own social media “timeline” or in his or her profile.

---

[183] WP 217 p. 36.

[184] WP 217 pp. 37–38.

[185] AG Opinion, footnotes 36 and 37; Request for a preliminary ruling by the Higher Regional Court of Düsseldorf regarding the case C-252/21, 24 March 2021, second question referred to the CJEU, p. 3. It should be noted, though, that the wording of the question does not reveal whether those data were actually collected and processed, or was it just an example made up by the referring court.

Third aspect in the assessment is the reasonable expectations of the data subjects regarding the use of the collected data, which is tied to the purpose limitation principle[186] – one of the key principles of the GDPR: data controllers cannot process data against the specified purposes of the data processing[187]. For example, if the purpose of the data collection from third party websites has been the personalization of advertising, disclosing the data publicly could be incompatible with that purpose. It is necessary to take into account the actual factual context rather than simply rely on text in small print.[188]

Last aspect in the assessment is the relationship between the data controller and the data subject. A notable imbalance between them may affect to the balancing test. In this case the data controller is a tech giant and data subjects are individuals, who may even be in vulnerable positions due to belonging to minority groups.

5.2.3.4 Additional safeguards

Lastly, when assessing all relevant aspects in the balancing test, the controller could apply additional measures in securing the balance between the legitimate interests of the controller and the fundamental rights and freedoms of the data subject, such as technical and organizational measures, anonymization or pseudonymization of the data[189], and measures to increase transparency.[190]

Returning to the German Facebook case and the assessment of the validity of the use of the legitimate interests justification, keeping in mind the balancing test as assessed above, it could be stated that whichever legitimate interest is assessed in this case, the balancing test triggers the validity of the legal basis. As Advocate General Rantos noted, it is difficult to see why it

---

[186] WP 217 p. 40.

[187] The purpose limitation principle is defined in the Article 5(1)(b) of the GDPR.

[188] WP 217 p. 40.

[189] Depending on the situation, anonymization or pseudonymization does not work every time. It has even suggested that anonymization and pseudonymization tools are not effective enough, and the reidentification is in many cases possible with data analytics tools. Read more about the problems of anonymization in Nadeza Purtova*, The law of everything. Broad concept of personal data and future of EU data protection law* (2018) 10 Law, Innovation and Technology 1.

[190] WP 217 p. 42.

would be necessary to process the data in question to pursue the legitimate interests of Meta, while the data originate from third-party sources.[191]

5.2.4 Effectiveness of a consent

The sixth question referred to the CJEU was about the effectiveness of a consent regarding the practice at issue, and the referring court was asking whether Articles 6(1)(a) and 9(2)(a) should be interpreted as meaning that "consent within the meaning of Article 4(11) of that regulation may be given effectively and freely to an undertaking having a dominant position in the national market for online social networks for private users"[192]. The Advocate General did not provide an unambiguous answer to the question, but rather stressed that the mere fact that the undertaking is in a dominant position (not necessarily in the meaning of Article 102 of the Treaty on the Functioning of the European Union (TFEU) but rather in the German competition law) does not lead to a clear imbalance between the data subject and the controller, as meant in the Recital 43 of the GDPR.[193]

The dominant position of the controller could still be a factor when assessing if data subjects can consent freely, creating such an imbalance between the controller and the data subject, and the validity of the consent shall be evaluated against this fact on case-by-case basis.[194] It remains to be seen whether the consent could be relied on, or whether it was valid or not.

However, according to the Advocate General, in (case-by-case) assessment of the validity and lawfulness of the consent, it should not be interpreted to be given freely and effectively, if the data subject did not have a genuine and free choice to consent, or he or she could not refuse or withdraw the consent without detriment to the service at question. Moreover, if the consent is a prerequisite to the performance of a contract and the consent would be required for processing activities that are not necessary for the performance of the contract, the consent

---

[191] Advocate General Opinion, paragraphs 64-66, p. 12.

[192] Request for a preliminary ruling of 24.3.2021 regarding the Case C-252/21, Higher Regional Court of Düsseldorf, 6th question, p. 5.

[193] Advocate General Opinion, paragraphs 75–77, p. 13.

[194] *Ibid*.

should be interpreted as invalid and not freely given, and a separate consent should be acquired to different processing purposes, instead of bundling the consent.[195]

Even though processing special categories of personal data (sensitive data) has been mostly left out of the scope of this paper, it should be noted that according to the AG Rantos, consent should not be sufficient to justify processing of sensitive personal data collected by cookies or similar technologies as meant in the ePricacy Directive[196], as it would not be sufficient to fulfil the requirements of an explicit consent as meant in Article 9(2) of the GDPR[197].

[195] Advocate General Opinion, paragraph 74, p. 13.

[196] Advocate General Opinion, paragraph 45, p. 9.

[197] Advocate General Opinion, footnote 59 p. 21.

6 Conclusions

6.1 Outer limits of data collection through sources other than data subjects

6.1.1 Legal bases for data collection and the principles data processing

Throughout this paper the legal bases for data collection and the principles of data minimization as well as purpose limitation has played key roles in defining the outer limits of the collection of personal data from sources other than the data subjects. The regulation concerning those aspects are backbones in the GDPR as well, forming a framework to protect the freedoms and reasonable expectations of the data subjects. To sum up, when collecting the data from third parties or other external sources the processing of data should be planned and organized in a timely manner to avoid confusions regarding the applicable and justified legal basis and the limits of processing. It must not be forgotten that each controller bears the burden of proof for lawfulness of their processing activities.

Records of processing activities, as meant in Article 30 of the GDPR, may help in defining the roles and responsibilities of all the key actors at the scene. Both the data collectors and the data recipients must be aware of their role and responsibilities towards data subjects and other parties in each activity.

First, when the consent of the data subject is relied upon as a legal basis for processing, particular attention must be paid to the purposes of the processing and the processing activities covered by the data subject's consent. Keeping the data subjects informed of all processing purposes and activities as well as all the parties taking part in the processing is essential to comply with the GDPR.

However, for AI development purposes keeping in mind the characteristics of Big Data, consent is a problematic legal basis due to its strict requirements regarding, among other things, obtaining consent, voluntariness and withdrawal of consent.

Second, relying on the legal basis of the performance of a contract should be limited to what is absolute necessary, without prejudice to the applicability of the legal basis in data collection from third parties. Consequently, this legal basis does not provide controllers in the field of Big Data, AI, or even social platforms, very trustworthy legal basis, as it already has been seen in this paper.

Third, legitimate interests of the controller or of a third party provides needed flexibility to the GDPR, although not without limitations in order to safeguard fundamental rights of data subjects. By implementing secure technical framework for data processing and informing the data subjects of the processing activities transparently, this legal basis may offer the best solution for AI developers as well.

Whichever is the legal basis to be applied, the principles of data minimization and purpose limitation should be kept in mind, especially in the field of AI, where the access to (open) data is of utmost importance.

6.1.2 Are there any lessons learned from the German Facebook case?

While most of all controllers aim to comply with the GDPR and by doing so making the online environment safer, some key actors, due to their reliance on data collection, do their best to avoid certain requirements and therefore collect extensive amounts of personal data in order to enable profile their customers as accurately as it is possible. The German Facebook case illustrated the problems stemming from the extensive data collection and the lack of adequate compliance measures within the organization.

While lacking the decision of the CJEU the case should be referred to narrowly, but same key issues arise form the case: none of the legal bases (should) work when the data is being collected through parties other than the data subject without their knowledge. The primary object of the GDPR is to enforce the rights to privacy instead of pushing organizations through unnecessary compliance measures. The case also illustrated how the data protection law is interconnected to the competition law.

6.2 Looking ahead: is the GDPR future-proof?

6.2.1 Big Data, AI, and the GDPR

While protecting the core of right to privacy the GDPR has its drawbacks as well especially to innovation and development. With the new rise of AI more focus has been put on privacy issues, highlighting both the importance of the core of the data protection and some issues the GDPR has not been able to tackle, or may even have unnecessarily hindered the innovation. In this chapter the extensive definition of personal data, as well as the core of the right to privacy are discussed in relation to the development of AI, among others.

AI has not been defined nor regulated comprehensively before. A proposition for regulation (Artificial Intelligence Act[198]) has been published by the EU. There has been a lot of discussion about the compatibility of the GDPR and AI and Big Data, envisaging the current and upcoming problems.[199] In this chapter the focus in put on the definition of the personal data and the principles of data minimization as well as purpose limitation in the light of AI development.

6.2.2 The definition of personal data

In her article, professor Nadezhda Purtova provocatively but aptly criticizes the EU law concept and definition of the personal data and envisages that even weather information could be defined as personal data.[200] Development of new AI applications and technologies need significant amounts of trained and untrained (raw) data and datasets, including information, that in broad interpretation of the term could be defined as personal data within the meaning of the GDPR. Data privacy law should keep sensitive information private, but not to hinder development of new innovations and technologies, unless the innovation in question is aimed to illegal or immoral purposes. Therefore, as Purtova has presented in her article, a new way of interpreting the core of data protection law should be introduced upon realization of big data applications.

With "onlife"[201] technologies it is possible to collect and store vast amounts of data, of which only a fraction would be relevant to protect, but as long as the data subject is identifiable from the data, the whole data would be defined as personal data and thus, the GDPR applies. As it

---

[198] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, COM/2021/206 final.

[199] See, for example: Martin Ebers, 'Regulating Al and Robotics: Ethical and Legal Challenges' in Martin Ebers and Susana Navas (eds), Algorithms and Law (CUP 2020); European Parliamentary Research Service, 'The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence' (2020); Lilian Mitrou, 'Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) "Artificial Intelligence-Proof'?' (2019); and Tal Z Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 996; Matthew Humerick, 'Taking Al Personally: Howthe EU Must Learn To Balance the Interests of Personal Data Privacy &Artificial Intelligence' (2018) 406-407.

[200] Nadeza Purtova, The law of everything. Broad concept of personal data and future of EU data protection law (2018) 10 Law, Innovation and Technology 1, pp. 57–59.

[201] Luciano Floridi came up with the term *onlife* when referring to a world where *online* and *offline* worlds or realities merge in a hyperconnected reality, that is "onlife". More of the term in Luciano Floridi, "Introduction" in Luciano Floridi (ed), The Online Manifesto. Being Human in a Hyperconnected Era, Springer 2015.

has been illustrated before, large datasets containing personal data hinder the development of AI models.

To train AI models would indeed be easier without the need to comply with the requirements of the GDPR. The application of the regulation could be avoided by claiming that the processing does not involve personal data, by first anonymizing or pseudonymizing the data, so that natural persons would not be directly or indirectly identifiable from datasets.[202] Due to challenges in trustworthiness regarding those techniques[203], this option is not preferable, even though there are limits to the reasonably likely means available to the controller to identify data subjects.[204]

Another not so widely discussed alternative to avoid the application of the GDPR altogether would be to process "transient data", which would not be defined as personal data and thus the GDPR would not apply.[205] This point of view could be criticized as not being technology-neutral by allowing some technologies to bypass the application of the GDPR and to avoid its requirements.

Could we still adopt a different approach here? More discussion would be needed of the definition of the personal data, as it covers basically anything that relates to a data subject. While ensuring the core of the data privacy, the definition could be narrowed down to only cover data that actually has value in the sense of privacy and right to private life. Those datapoints could be address details, contract information, credit scores and information on income and taxation, health information etc. From the point of view of fundamental rights, I don't really see the point in protecting the information on cars, clothes, sports or any other datapoints laying outside the core of the data protection.[206] The list of protectable datapoints

---

[202] Manon Oostveen, Identifiability and the applicability of data protection to big data, International Data Privacy Law, Volume 6, Issue 4, November 2016, pp. 299–309.

[203] Rocher, Hendrickx and de Montjoye, 'Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models' (2019) 10 Nature Communications 3069.

[204] CJEU, Case C-582/14, Breyer, paras. 45–49.

[205] George, Damian, and Reutimann, Kento. 'GDPR Bypass by Design? Transient Processing of Data under the GDPR'. International Data Privacy Law 9, no. 4 (2019): 14.

[206] See also Purtova 2018 p. 42. It is worth noticing that nor Purtova is or am I suggesting a narrower scope of data protection law, but instead a narrower scope of the definition of the personal data in order to foster the core of the protection.

could even be regulated through a decision of the EU Commission, or even through national laws, enforcing the principle of subsidiarity in the EU.

With the narrower definition of the personal data the compliance costs and hindrances of the GDPR could decrease. The phenomenon is no doubt that simple, but a lot more discussion is needed in this area to proactively shape the European Single Market in a direction that is both safe for individual data subjects as well as an encouraging environment for startups and businesses to develop the technology of our future.

6.2.3 Protecting individuals by hindering markets? Principles of data minimization and purpose limitation

In addition to the definition of the personal data, the principles of data minimization and purpose limitation as meant in the GDPR add some friction to innovation and development of AI. According to the principle of data minimization the amount and nature of personal data being collected should be narrowed to what is necessary. While Big Data, as is visible from the term, and consequently AI as well, relies on extensive datasets, this principle has become a hindrance in development purposes[207], while at the same time big data analytics companies may not comply with the principle quite well[208]. The principle of data minimization (among other requirements) requires the controllers to define the purposes of the processing and especially the datapoints to be collected upon the collection at the latest.

For AI development the principle is not compatible with the functioning of the development activities: AI is trained with large datasets, and the algorithm may connect dots in a way that is beyond human understanding[209], and therefore knowing which datapoints to collect, for which purposes and with which legal basis may be a big challenge for AI developers.[210]

---

[207] Michele Finch and Asia Biega, 'Reviving Purpose Limitation and Data Minimisation in Personalisation, Profiling and Decision-Making Systems' (2021) Max Planck Institute for Innovation and Competition (research paper series) pp. 30–31.

[208] Zarsky 2017 pp. 1010–1011.

[209] The phenomenon is also known as a black box effect. See, for example, The impact of the general data protection regulation on artificial intelligence", European Parliament. Directorate General for Parliamentary Research Services, 2020, p. 14.

[210] See, for example, Paloma Kroot Tupay, Martin Ebers, Jakob Juksaar & Kea Kohv, 'Is European Data Protection Toxic for Innovative AI? An Estonia Perspective' (2021) 30 Juridica Int'l 99. p. 104.

Principle of purpose limitation has basically the same problems, requiring the controllers to beforehand planning which does not work very well with the AI development. Some flexibility is offered to the principle, however, being that new purposes are allowed insofar as they are not incompatible with the original purposes, according to Article 6(4) of the GDPR, taking into account the reasonable expectations of the data subjects as well as the consequences of the processing in light of new purposes.[211] Avoiding the incompliance with the purpose limitation principle by defining one of the purposes as, for example, "Research and development purposes" or similar kind of open-ended and unclear statement is not possible, nor should not be possible.[212] Still, when collecting (or scraping) the data to develop AI models the principle brings obstacles.

In the limits of this paper, I do not have a possibility to introduce a better framework to principles of the GDPR, but much more political and research-related discussion is needed in this area as well. An alternative solution has been implemented in California with the new California Consumer Privacy Act (CCPA), introducing a "do not sell my data" opt-out option for individuals not wanting to take part in extensive data collection.[213] Keeping in mind the location of Silicon Valley in California, the home base of technology conglomerates, the data privacy legislation, to this date, still has not scared those away from the area. I think we still may have something to learn in Europe on balancing the rights of consumers and the needs of businesses. During my lifetime I would like to see legislation that is able to protect the hard core of the data privacy while at the same time safeguarding the innovation and technological development in the EU.

---

[211] Tupay – Ebers – Kohv 2021 p. 103.

[212] Art 29 WP, 'Guidelines on Transparency under Regulation 2016/679' (2017) p. 9.

[213] For a brief introduction, see the webpage of State of California Department of Justice at https://oag.ca.gov/privacy/ccpa (Accessed April 26, 2023).