To Scrape or Not to Scrape? The Lawfulness of Social Media Crawling under the GDPR[1]

C. Altobelli, N. Forgó, E. Johnson & A. Napieralski[2]

[1] The authors are (partly) funded by the CREST and CONNEXIONs Projects. These project have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No. 833464 (CREST) and No. 786731 (CONNEXIONs).

[2] University of Vienna, Faculty of Law, Department of Innovation and Digitalisation in Law.

Introduction

Data from past events, real-time data and news data are all present on social media. As such, the personal data generated from social media use is vast, detailed and many times sensitive. Consequently, third parties and the platforms themselves can gain valuable insights from the data. Given the accessibility and availability of much of the social media data, the information is increasingly attractive to researchers and other third parties alike.

Personal data on social media exists in different forms of availability. This chapter focuses on social media data which is publically available and the data collection of which is regulated by the General Data Protection Regulation (GDPR).[3] This data collection context contrasts to that of privately held personal data, which requires police and other competent authorities to request access to the data from the online service provider. In the latter instance, Directive 2016/680 regulates data processing and will not be applicable to this chapter.[4]

The following chapter will describe the different categories of personal data available on social media platforms and broadly outline the techniques employed to gather social media data. Following this, the chapter will examine the legal bases for the further processing of social media data by third parties generally and then more specifically for researchers.

The legality of third party processing of social media data has received significant attention in recent years. An obvious example is the Facebook and Cambridge Analytica case which came to fruition when Facebook was exposed for providing the political consulting firm, Cambridge Analytica, with access to the personal data of over 50 million Facebook users.[5] Further, a 2018 study found that 60% of Twitter users surveyed were unaware that publically available Tweets

[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] Directive (EU) 2016/680 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 by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of such Data, and Repealing Council Framework Decision 2008/977/JHA.

[5] Tim Adams, 'Facebook's Week of Shame: the Cambridge Analytica Fallout' (*The Guardian*, 24 March 2018) <https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence> accessed 11 December 2019.

can be used for research.[6] Hence, there appears to be a disparity between the knowledge and expectations of the platform users and what is legally permitted in third party further processing. This chapter aims to clarify the legal environment in the EU surrounding the further processing of social media data dealing with issues of data protection only (ignoring, therefore, other legal issues such as copyright, access rights etc.).

1. Type and Formats of Personal Data on Social Media

Social media consists of interactive platforms for user generated data dissemination. At the centre of social media is the facilitation of the creation and sharing of ideas and data, as well as the ability to represent the human self. The social media data can take various forms, whether text, photographs, voice recordings or videos. Depending upon the content of a data set, it can, according to the GDPR come under the special category of personal data. The GDPR states that "personal data means any information relating to an identified or identifiable natural person."[7] This chapter focuses on the lawfulness of social media data crawling in accordance with the GDPR.

Web scraping is a form of data extraction that combines "contents of interest from the Web in a systematic way".[8] "A web scraper is comprised of two parts, the one is the crawler and the second is data extractor".[9] Data scraping is a form of copying data and storing it on a central database for future analysis. Web crawling is the first part of the online data extraction process. The crawler, which is also known as a web robot, or bot, "systematically scans the World Wide Web" with the aim of collecting data. The scraping process can involve the use of APIs (Application Programming Interface), whether third party APIs or those offered by the platform. APIs are a set of tools and protocols that allow programmers to build software applications. More specifically, APIs "expose services or data by a software application through a set of predefined resources, such as methods, objects, or URIs (Uniform Resource Identifier). By using these resources, other applications can access the data or services without having to

[6] Casey Fiesler and Nicholas Proferes, "'Participant" Perceptions of Twitter Research Ethics' [2018] Social Media + Society 1-14.

[7] GDPR, Article 6(1).

[8] Glez-Pena and others, 'Web scraping Technologies in an API World' [2014] 15(5) Briefings in Bioinformatics 788-797.

[9] M. Saniya Parvez and others, 'Analysis of Different Web Data Extraction Techniques' [2018] International Conference on Smart City and Engineering Technology.

implement the underlying objects and procedure."[10] APIs can create programs to crawl social media. In referring back to the GDPR definition of processing, Article 4(2) encompasses automated processing. Moreover, as online data scraping and crawling comprises of several of the actions listed in the Article 4(2) definition including collection, organisation, storage etc., then the GDPR regulates the processing social media personal data in this context.

Article 6 GDPR provides the possible legal bases that allow for the processing of personal data. On a social media platform, this can include a person's name, email address, age, occupation, a photograph of them and so on. However, these examples do not cover the entirety of social media data. There exist more sensitive forms of personal data on social media platforms and elsewhere which Article 9(1) GDPR titles 'special categories of personal data'. Such data is data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.[11] When it comes to text data, Tweets or Facebook posts can (and frequently do) consist of political opinions and religious and philosophical beliefs.[12] Equally, information sections of social media sites can indicate an individual's sexual orientation. For example, the Facebook Details section allows a user to input their relationship status. This information can be made publically available by the user in the Facebook user settings. Similarly, photographs and posts can also provide data concerning a person's sex life and sexual orientation.

Article 9 also encompasses biometric data. 'Biometric data' is defined in the GDPR as personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data. While biometric data consists of any data which uniquely identifies an individual and can range from finger prints, iris scans and even vein recognition, the most commonly found biometric data set on social media

[10] Michael Meng and others, 'Application Programming Interface Documentation: What Do Software Developers Want?' [2018] 48(3) Journal of Technical Writing and Communication 295-330.

[11] GDPR, Article 9(1).

[12] Gaby Hinsliff, 'Trash Talk: How Twitter is Shaping the New Politics' (*The Guardian*, 31 July 2016) <https://www.theguardian.com/technology/2016/jul/31/trash-talk-how-twitter-is-shaping-the-new-politics> accessed 11 December 2019.

platforms is facial images. Social media contains masses of facial images, for example, in 2014 Facebook research on the DeepFace facial recognition technology was able to train DeepFace on the "largest facial dataset to-date".[13] Shift to June 2019 and Facebook had on average 1.59 billon daily active users. If only half of these daily users have profile pictures of their faces, have posted pictures online, or have even been tagged in photographs by friends, that is an alarming number of faces on only one social media platform. The question of whether facial images on social media platforms are biometric data, and therefore a 'special category of personal data' is determined by the way in which the data is processed. Specifically, Recital 51 GDPR states that the processing of photographs should not systematically be considered to be processing of special categories of personal data as they are covered by the definition of biometric data only when processed through a specific technical means allowing the unique identification or authentication of a natural person. Therefore, a photograph of a face is not *per se* biometric data, though Facebook's automatic image tagging technology would make the facial images biometric data due to the technical means used to employ the unique identification of its users.[14]

Regardless of whether a photograph can be considered biometric data as set down by the GDPR, one must consider whether a photograph contains data concerning health. Article 4(15) GDPR states that 'data concerning health' means "personal data related to the physical or mental health of a natural person". The diagnosis of many medical conditions feature elements of visual patient observation. For example, a photograph can demonstrate skin conditions, physical disabilities, visual impairment, and pregnancy, among a vast number of other medical conditions. Further, the GDPR does not distinguish between the data type containing health data, nor does it distinguish between physical and mental health data.[15] As such, a user post indicting their mental health status, depending on the specific situation could be data concerning health. Therefore, a photograph or piece of text on a social media site can, on a case by case basis, contain 'data concerning health'.

---

[13] Yaniv Taigman and others, 'DeepFace: Closing the Gap to Human-Level Performance in Face Verification' [2014] Facebook Research, Conference on Computer Vision and Pattern Recognition (CVPR).

[14] Facebook, 'What is the Face Recognition Setting on Facebook and How Does it Work?' (*Facebook*) <https://www.facebook.com/help/122175507864081> accessed 14 February 2020; in this case Facebook currently requires users to turn this feature on. It may be argued, depending on the means by which permission is gathered from the data subject, Facebook may have obtained the required explicit consent for processing under Article 9(2)(a) GDPR.

[15] GDPR, Recital 35.

Additionally, facial images, and other personal images may also reveal the racial and/or ethnic origins of an individual. However, these categories, as may also be the case with health data, may also not accurately reveal such personal details. These representations may indicate a specific special category of data, but given the limitations of image data, they may also be incorrect. The presence of additional data which can also found on social media may contribute to the provision of a confirmation of the presence of the special category of personal data. However, examining the data sets alone may not definitively confirm the presence of particular categories of personal data. As such, the question of whether the data concerned is a special category should be considered objectively and with consideration to additionally available data.

These examples of the processing of special categories of personal data on social media platforms are demonstrative of the vastly different types and formats of data existing on social media platforms. A further complicating factor is the notion of ambiguous data sets regarding whether they would fall under Article 6 or Article 9 GDPR, or contain personal data, which could come under a combination of the two.

Evidently, the type and categories of personal data are not straightforward and can potentially induce a legal minefield for controllers and processors.

2. Personal Data Collection Techniques on Social Media Platforms

The particular data processing methods applied to data collection on social media often depend upon the aim to be achieved, the type of data intended to be collected and the availability of the data. For GDPR regulated data collection online, without the presence of the consent, and with regard to the privacy expectations of the data subject, only publically available personal data can be legitimately processed by third parties. As such, blog sites including Twitter are obvious targets. Further examples include Facebook, Instagram and LinkedIn, which also all provide publically available personal data based on user privacy preferences.

Three primary steps for data collection online have been identified.[16] Firstly, the action of finding, which uses search tools to identify target information on the web through the use of search engines for example or tools. Then the extraction, a process which employs tools to

[16] Ray Poynter, *The Handbook of Online and Social Media Research: Tools and Techniques for Market Researchers* (1st edn, John Wiley & Sons 2012) 223.

"extract the information found and store it in an accessible form, for example using web scrapers".[17] Scraping consists of "collecting online data from social media and other Web sites in the form of unstructured text".[18] Automated web scraping allows for the extraction of large amounts of data, effectively emulating "the behaviour of people using the web", but on a much larger and faster rate allowing for the selection of specific search terms.[19] The final step is the analysing which uses software such as Leximancer[20] to analyse and integrate large sources of data enabling users to derive insight. Search tools may be provided by the platforms themselves such as with Twitter's Advanced Search[21] or alternatively with web-based clients such as TwitScoop.[22] Google Analytics[23] is a freely accessible source, which allows users to view website activity including through real-time analytics.

Regardless of the specific service, web scraping enables researchers to collect large amounts of personal data on a single individual and on large target groups. The following section will examine the current legislative approaches to the collection of social media data for research purposes and for non-research purposes.

3. Different legislative approaches for processing for research purposes

A 'purpose' in this context is the reason something is done,[24] and namely why the processing of personal data is done. Processing for research purposes uses the data and/or information, which can be derived from the data to contribute to the gaining and expansion of knowledge. This activity is, in part, defined by the aim to be achieved.

---

[17] Ibid.

[18] Bogdan Batrinca and Philip C. Treleaven, 'Social Media Analytics: a Survey of Techniques, Tools and Platforms' [2015] 30(1) AI & Society 89-116.

[19] Ray Poynter, *The Handbook of Online and Social Media Research: Tools and Techniques for Market Researchers* (1st edn, John Wiley & Sons 2012) 223 at 232.

[20] Leximancer, 'Text In Insight Out' (*Leximancer*) <https://info.leximancer.com/> accessed 11 December 2019.

[21] Twitter, 'Advanced Search' (*Twitter*) <https://twitter.com/search-advanced?lang=en> accessed 11 December 2019.

[22] Twitscoop, 'Worldwide Twitter Trends' (*Twitscoop*) <https://www.twitscoop.com/> accessed 24 February 2020

[23] Google, 'Analytics' (*Google*) <https://analytics.google.com/analytics/web/provision/#/provision> accessed 24 February 2020.

[24] Cambridge Dictionary, 'Purpose' (*Cambridge Dictionary*) <https://dictionary.cambridge.org/dictionary/english/purpose> accessed 11 December 2019.

Distinctive legislative approaches to research purposes are demonstrated in EU legislation. Article 179 of the Treaty of the Functioning of the European Union (TFEU),[25] which establishes the European research area states that:

> "1. The Union shall have the objective of strengthening its scientific and technological bases by achieving a European research area in which researchers, scientific knowledge and technology circulate freely, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties."[26]

The TFEU lays the foundations of the Member States as a collective Union. Thus, at the one of the bases of the EU, a legislative statement to encourage the European research area is affirmed.[27] This affirmation is made separate to that of operating for commercial purposes.

The GDPR reflects this legislative distinction between processing for research purposes, and then processing for other purposes. While the GDPR specifically names 'scientific' and 'historic research purposes', it does not define them. Recital 159 GDPR states that "[f]or the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner". The Article 29 Working Party is of the opinion that the term 'scientific research' "may not be stretched beyond its common meaning" and that scientific research "in this context means a research project set up in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice".[28] Therefore, processing for research purposes is not only defined by the end goal(s) of the research, but also the context and the methodology applied.[29]

The GDPR distinguishes processing for research purposes from other forms of processing. It has been remarked that, in comparison with the GDPR's predecessor, Directive 95/46/EC, the

---

[25] Consolidated Version of the Treaty on the Functioning of the European Union.

[26] Treaty of the Functioning of the European Union, Article 179.

[27] Treaty on the Functioning of the European Union, Preamble.

[28] Article 29 Working Party, *Guidelines on Consent under Regulation 2016/679* (WP259, 2017) at 28.

[29] *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities* (OECD Publishing, 2015) at 45. The Frascati Manual, the international standard for the use of research and experimental development statistics (R&D), requires the satisfaction of five core criteria: novel, creative, uncertain, systematic, transferable and/or reproducible. R&D must also be made up of basic research, applied research and experimental development.

GDPR is more forthcoming toward research through the inclusion of the Article 89(2) derogations.[30] One commentator on the topic stated that the purpose of the Article 89 research derogations is to "instantiate into law what is already good scientific practice".[31] The GDPR contains other provisions that appear to be favourable towards research, namely Article 5(1)(b) which asserts that further processing for "historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes".[32] Therefore, the GDPR provides unique provisions for processing for research purposes and then generic all other processing activities.[33]

These provisions will be discussed in more detail further in this chapter, but for now, their presence in the GDPR demonstrates the different legislative approaches towards the two different situations.

4. Lawfulness of Social Media Crawling

The GDPR explicitly distinguishes between processing of personal data for purposes for which they were initially collected and processing activities for further purposes.[34] To simplify this distinction this chapter refers to the former as 'primary processing' and the latter as 'further processing.' Although 'further processing' is not explicitly defined in the text of the GDPR, the Article 29 Working Party referred to it as "the use of personal data that was originally collected for something else."[35] Considering social media crawling, on one hand, the primary processing activity can be generally identified as the processing carried out by the social media platform itself, acting as the primary controller collecting the personal data. On the other, when a different controller carries out crawling on the personal data present on the platform obtained from APIs for a different purpose than the original one, this would generally qualify as further

[30] Chih-Hsing Ho, 'Challenges of the EU General Data Protection Regulation for Biobanking and Scientific Research' [2015] 25 Journal of Law, Information and Science 1, 84-103.

[31] Edward S. Dove, 'The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era' [2018] 46 The Journal of Law, Medicine & Ethics 4, 1013-1030.

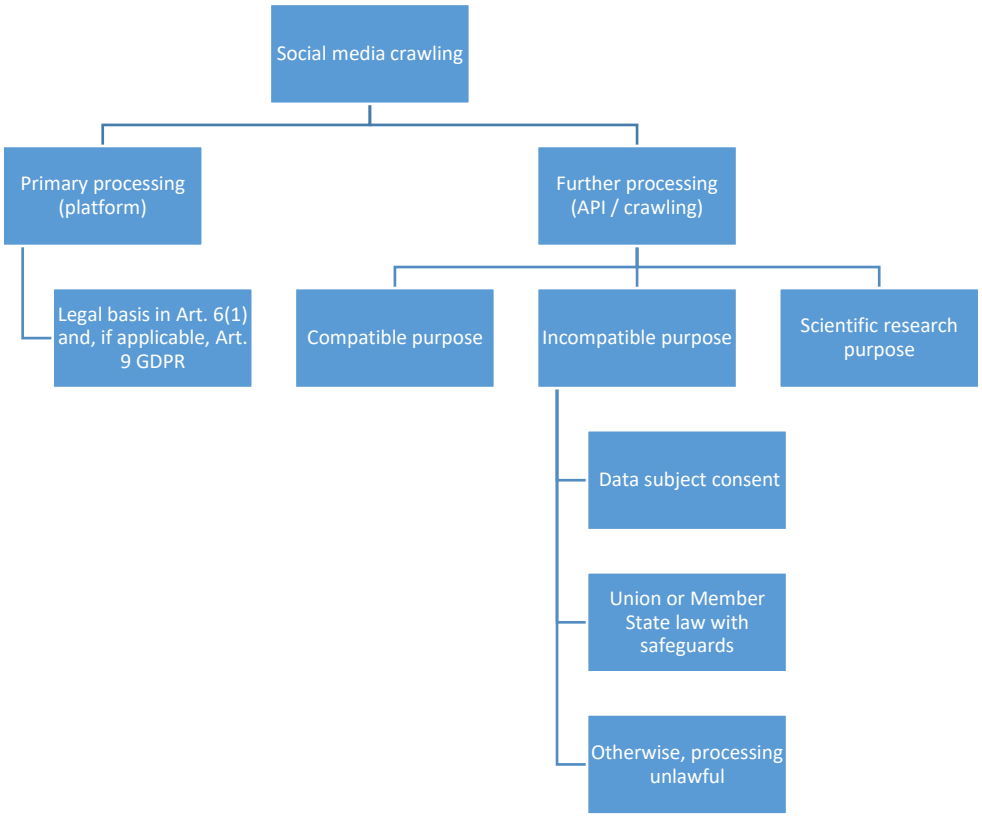
[32] GDPR, Article 5(1)(b).

[33] Excluding those purposes listed in Article 2(2) GDPR.

[34] GDPR, Recital 50.

[35] Article 29 Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679* (WP 251) at 18.

processing.[36] It follows that the controller of the primary processing activities must comply with Article 6(1) and 9(2) GDPR, whereas the secondary controller of the further processing must fulfil the conditions of Article 6(4) GDPR. Such distinction is important to safeguard the interests of the data subject. In fact, as discussed in depth in the following sections, the grounds for further processing can be considered limited as opposed to the general provision applicable to the primary processing. Thus, the limited grounds allowing further processing ensure that, once the data is collected, the data subject is capable of exercising some control on further processing activities. In particular, it restricts controllers from exploiting collected data through a "chain of processing activities", where the purposes of processing are substantially different from the original ones. The following diagram synthesises the steps required to lawfully carry out social media crawling in accordance with the GDPR:



[36] However, some authors analyse mining activities on publicly available data as a new processing activity. See, for example: Elena Gil González and Paul de Hert, 'Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data – An Analysis of GDPR Provisions and Principles' [2019] 19 ERA Forum 4; Benjamin Bremert, 'Legal Aspects of Text Mining Publicly Available Data' [2017] ULD, Independent Centre for Privacy Protection.

4.1. Lawfulness of Primary Processing by Social Media Platform

As mentioned in the previous section, in this scenario, crawling activities carried out by a controller different from the social media platform itself generally fall within the meaning of further processing. However, it is important to briefly discuss the issue of lawfulness of the primary processing, which may as well include crawling activities. Nonetheless, this section does not intend to be an exhaustive description of lawfulness of social media processing activities. According to Article 5(1)(a) GDPR, the processing of personal data must always be lawful, fair and transparent. In order for the primary processing activity to be lawful one of the legal grounds in Article 6(1) GDPR must be applicable.

4.1.1 Consent

When considering social media processing activities, the primary legal basis is consent by the data subject.[37] It is often the case that data privacy policies of social media platforms specifically reserve the right to share the data subject's data to third parties.[38] For example, according to Twitter's privacy policy, by publicly posting content by 'Tweeting', the data subject is "directing us to disclose that information as broadly as possible, including through our APIs, and directing those accessing the information through our APIs to do the same."[39] The GDPR defines lawful consent as a freely given, specific, informed and unambiguous indication of the data subject's wishes.[40] A question that is beyond the scope of this chapter, but still very relevant, is whether privacy clauses within the terms of services of a social media platform can be deemed in line with the requirements of consent embedded in the GDPR.[41] However, the

[37] GDPR, Article 6(1)(a).

[38] Benjamin Bremert, 'Legal Aspects of Text Mining Publicly Available Data' [2017] ULD, Independent Centre for Privacy Protection at 4.

[39] Twitter, 'Privacy Policy' (*Twitter*, 2019) <https://twitter.com/en/privacy> accessed 12 December 2019.

[40] GDPR, Article 4(11) and Article 7.

[41] For further literature on consent see, for example: Eleni Kosta, 'Consent in European Data Protection Law' [2013] Martinus Nijhoff Publishers; Benjamin Bergemann, 'The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection' in Marit Hansen, Eleni Kosta, Igor Nai-Fovino and Simone Fischer-Hübner (eds), *Privacy and Identity Management. The Smart Revolution* (Springer 2018); Bart W. Schermer, Bart Custers and Simone van der Hof, 'The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection' [2014] 16 Ethics and Information Technology; Spyros Polykalas, 'Assessing General Data Protection Regulation for Personal Data: is the End of "Take it or Leave it" Approach for Dowloading Apps?' [2017] The Seventh International Conference on Social Media Technologies, Communication, and Informatics.

discussion in this chapter lies on the presumption that, if applicable, the original consent obtained by the social media platform is valid.

4.1.2 Performance of a contract

Another possible legal basis for the primary processing of social media data is when the processing is 'necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contact'.[42] The GDPR provides little additional explanation for this provision other than that processing is lawful in this context only when there is an element of necessity.[43] Given the ease and availability of the collection and processing of personal data online, the Article 29 Working Party asserted that the purpose of data collection must be 'clearly and specifically identified' and as such, 'a purpose that is vague or general, such as for instance 'improving users' experience', 'marketing purposes', 'IT-security purposes' or 'future research' will-without more detail, usually not meet the criteria of being 'specific''.[44] In this instance, while contracts relating to online services are not usually completed on an individual basis, both the purpose limitation and the data minimisation principles must be taken into consideration.[45]

4.1.3 Legitimate interests

A further legal basis that is relevant for the primary processing activity on social media is when the processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party.[46] In order to evaluate whether the interests of the controller are legitimate, a three-step test must be fulfilled, ensuring that any processing activity falling outside of the scope of the other legal basis meet specific requirements.[47] Firstly, the controller must pursue a legitimate interest, which refers to the broad objective the controller aims to achieve with the processing.[48] As regards processing activities on social media, the controller may rely on an

---

[42] GDPR, Article 6(1)(b).

[43] GDPR, Recital 44.

[44] Article 29 Working Party, *Opinion 03/2013 on Purpose Limitation* (WP203 2013) at 15–16.

[45] 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* (EDPB Guidelines, 2019).

[46] GDPR, Article 6(1)(f).

[47] 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* (WP 217) at 9.

[48] Ibid at 23.

interest in profiling a subject with the objective of, for example, targeting it with advertisement.[49] In this regard, Recital 47 GDPR explicitly mentions that "the processing of personal data for direct marketing purposes may be regarded as carried out for a legitimate interest."[50] Secondly, the data processing activity must be necessary to achieve the legitimate interest.[51] This test involves assessing whether there is a less intrusive way upon the subjects' rights to achieve the same interest.[52] When considering a social media platform, the controller should limit reliance on this basis only for that data required for the specific processing activity.[53] Finally, a balancing test must be carried out between the legitimate interests of the data controller and the interests or rights and freedoms of the data subject.[54] It follows that, after having identified the interests in favour or against the data processing, these interests must be weighed against each other.[55] In this regard, and particularly relevant to social media crawling, Recital 47 emphasises that "the interests and fundamental rights of the data subject could in particular override the interest of the data controller where personal data are processed in circumstances where data subjects do not reasonably expect further processing."[56]

In practice, if one takes the Data Policy of Facebook as an example, the platform relies on the ground of legitimate interests for a number of purposes.[57] Among others, it claims it has legitimate interests in line with Article 6(1)(f) GDPR "for providing measurement, analytics and other business services where we are processing data as a controller."[58] According to its

---

[49] Elena Gil González and Paul de Hert, 'Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data – An Analysis of GDPR Provisions and Principles' [2019] 19 ERA Forum 4 at 606.

[50] GDPR, Recital 47.

[51] 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* (WP 217) at 29.

[52] Elena Gil González and Paul de Hert, 'Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data – An Analysis of GDPR Provisions and Principles' [2019] 19 *ERA Forum* 4 at 606.

[53] Benjamin Bremert, 'Legal Aspects of Text Mining Publicly Available Data' [2017] ULD, Independent Centre for Privacy Protection at 7.

[54] 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* (WP 217) at 33.

[55] Benjamin Bremert, 'Legal Aspects of Text Mining Publicly Available Data' [2017] ULD, Independent Centre for Privacy Protection at 7.

[56] GDPR, Recital 47.

[57] Facebook, Data Policy (*Facebook*, 2019) <https://www.facebook.com/about/privacy/legal_bases> accessed 12 December 2019.

[58] Ibid.

policy, the legitimate interests are "in the interests of advertisers, developers and other partners to help them understand their customers and improve their business (…)."[59] The policy specifically states that the users' interests or fundamental rights and freedoms do not outweigh their legitimate interests.[60] Hence, together with consent, social media platforms are likely to rely on legitimate interests as the legal basis for their data processing activities.

4.1.4 Special Category of Personal Data

As described previously, a vast portion of personal data on social media falls within the definition of "special category of personal data" under Article 9(1) GDPR. It follows that to circumvent the prohibition to process this type of data, the controller must fulfil one of the requirements in Article 9(2) GDPR. However, it is important to emphasise that even in case of special categories of personal data, one of the legal bases in Article 6(1) GDPR must apply, as Article 9(2) merely complements it with further safeguards.[61] Thus, in other words, it may be argued that Article 9(2) GDPR applies in addition to the general rules on data processing in Article 6 GDPR, as the former cannot be deemed a *lex specialis*.[62] This finding is confirmed in Recital 51 GDPR, which states that "in addition to the specific requirements for such processing, the general principles and other rules of this Regulation should apply, *in particular as regards the conditions for lawful processing*" (emphasis added). It is important to highlight that this issue is fiercely debated, with a number of data protection scholars arguing to the contrary.[63] As stressed by the Article 29 Working party, although some of the exceptions for

---

[59] Ibid.

[60] Ibid.

[61] Article 29 Data Protection Working Party, *Advice Paper on Special Categories of Data ("Sensitive Data")* at 5; European Data Protection Board, *Guidelines 3/2019 on Processing of Personal Data through Video Devices* v2.0, (EDPB Guidelines, 2020) at 68.

[62] This view is supported by many authors: Thilo Weichert, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Jürgen Kühling and Benedikt Buchner (eds), Datenschutz-Grundverordnung/BDSG (Beck 2018) at 4; Benjamin Bremert, 'Legal Aspects of Text Mining Publicly Available Data' [2017] ULD, Independent Centre for Privacy Protection at 7; Thomas Petri, 'Artikel 9' in Spiros Simitis, Gerrit Hornung and Indra Spiecker (eds), *Datenschutzrecht DSGVO mit BDSG* (Nomos 2019) at 2.

[63] See, for example: Eike Michael Frenzel, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Boris Paal and Daniel Pauly (eds*), Datenschutz-Grundverordnung Bundesdatenschutzgesetz* (Beck 2018) at 18; Holger Greve, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Martin Eßer, Philipp Kramer and Kai von Lewinski (eds), *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz und Nebengesetze* (Heymanns 2018) at 16; Peter Schantz, ‚Zulässigkeit der Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Peter Schantz and Heinrich Amadeus Wolff (eds), *Das Neue Datenschutzrecht*

processing special category of personal data have stricter requirements than those in Article 6(1) GDPR, such as the 'explicit' consent in Article 9(2)(a) GDPR, this does not apply to all the provisions.[64] In fact, Article 9(2)(e) GDPR, which foresees that controllers may process data which are manifestly made public by the data subject, appears more lenient than the conditions in Article 6(1) GDPR.[65] Thus, if the data subject manifestly makes personal data public, in most cases, the controller will have to rely on Article 6(1)(f) and fulfil the legitimate interest test described in the previous section.

Concerning processing activities relating to special category of personal data by a social media platform, the most relevant exceptions to the prohibition of processing are explicit consent and data manifestly made public by the data subject.[66] Under Article 9(2)(a) GDPR, the prohibition of processing is lifted when the data subject has given "an express statement of consent."[67] For example, in a social media setting this requirement can be met when the data subject is "able to issue the required statement by filling an electronic form."[68] As regards Article 9(2)(e), the GDPR is silent in defining the term 'made public'. Nonetheless, a number of scholars recognise that data is to be deemed as made available to the public when an indeterminate number of users can easily access it.[69] Therefore, on social media, whether specific data has been manifestly made public by the data subject will depend on whether only a limited number of users can access the information or whether it is accessible to the wider public.[70] Moreover, it is crucial that the personal data is deliberately made available by the data subject and not a third party.[71]

(Beck 2017) at 705; David Kampert, ‚Artikel 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Gernot Sydow (ed), *Europäische Datenschutzgrund-verordnung* (Nomos 2018) at 63.

[64] 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* (WP 217) at 14.

[65] Ibid at 15.

[66] GDPR, Article 9(2)(a) and 9(2)(e).

[67] Article 29 Working Party, *Guidelines on Consent under Regulation 2016/679* (WP 259, 2016) at 18.

[68] Ibid.

[69] Sebastian Schulz, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Peter Gola (ed), *Datenschutz-Grundverordnung* (Beck 2018) at 26; David Kampert, 'Artikel 9 Verarbeitung Besonderer Kategorien Personenbezogener' Daten in Gernot Sydow (ed), *Europäische Datenschutzgrund-verordnung* (Nomos 2018) at 31.

[70] Thilo Weichert, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Jürgen Kühling and Benedikt Buchner (eds), Datenschutz-Grundverordnung/BDSG (Beck 2018) at 82.

[71] Thomas Petri, 'Artikel 9' in Spiros Simitis, Gerrit Hornung and Indra Spiecker (eds), *Datenschutzrecht DSGVO mit BDSG* (Nomos 2019) at 57.

It follows that it must be clear for a user that, when visiting a social media platform, a specific account is linked to the data subject and the subject has manifestly published the data at stake.[72] However, in case of uncertainty, Article 9(2)(e) GDPR will not be applicable.[73]

4.1.5. Outcome of Lawfulness Analysis

The previous sections discussed the question of lawfulness of the primary processing by the social media platform, which may as well include itself crawling activities. Such analysis is crucial when determining whether scraping activities by a third party may be deemed to be in line with Article 6(4) GDPR, as further processing. After focusing on various legal basis in Article 6(1) GDPR which may be applicable to data processing on social media platforms, it can be concluded that such activities will in most cases fulfil the principle of lawfulness. In fact, either on the basis of consent, performance of a contract, or legitimate interests, the platform may scrape and analyse users' content independently from it being an entity with commercial interests. Nonetheless, as emphasised previously, the specific requirements of the specific legal basis must be fulfilled and lawfulness will depend on compliance with the other principles of data protection, such as data minimisation and purpose limitation.

4.2 Lawfulness of Social Media Crawling as Further Processing

The previous sections discussed the lawfulness of processing activities on social media by the platform itself. This section builds upon those findings to determine under what conditions social media crawling, as further processing, is in line with the GDPR. Lawfulness of further processing under the GDPR is strictly linked to the principle of purpose limitation. Indeed, the principle of purpose limitation determines that, firstly, personal data must be collected for specified, explicit and legitimate purposes and, secondly, it must not be further processed in a manner that is incompatible with those purposes.[74] The GDPR does not impose an obligation of compatibility, rather it prohibits incompatibility, as it uses a double negation rather than a

[72] Benjamin Bremert, 'Legal Aspects of Text Mining Publicly Available Data' [2017] ULD, Independent Centre for Privacy Protection at 7.

[73] Thilo Weichert, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Jürgen Kühling and Benedikt Buchner (eds), Datenschutz-Grundverordnung/BDSG (Beck 2018) at 80.

[74] GDPR, Article 5(1)(b).

positive statement: "further processing is authorised as long as it is not incompatible".[75] According to the Article 29 Working Party, such wording seems to highlight that "the legislators intended to give some flexibility with regard to further use."[76]

The principle of purpose limitation includes two elements.[77] The first, namely "personal data must be collected for specified, explicit and legitimate purposes" (purpose specification component) does not allow derogations.[78] Whereas the second, "it must not be further processed in a manner that is incompatible with those purposes" (use limitation component) allows a number of exceptions.[79] As regards the 'purpose specification component', no derogations are allowed because of its connection to the requirement of foreseeability under Article 8(2) of the EU Charter of Fundamental Rights.[80] Whereas, concerning the use limitation component, the GDPR provides precise derogations.[81]

The following derogations are envisaged in the GDPR:

i. Data is re-used for incompatible purposes with the consent of the data subject;[82]
ii. Data is re-used for incompatible purposes but on the basis of a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1) GDPR.[83]

Therefore, according to Article 6(4) GDPR, further processing is lawful if either the controller asks for the data subject's consent on the new purpose, there is a specific ground in European or national law (with safeguards), or the new purpose is deemed compatible with the original purpose (test of compatibility), as discussed later. In case of statistical, scientific and historic research purposes, further processing shall not be considered incompatible with the initial

[75] Wouter Seinen, Andre Walter and Sari van Grondelle, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' in Manel Medina, Andreas Mitrakas and others (eds), *Privacy Technology and Policy* (Springer 2018) at 155.

[76] Article 29 Working Party, *Opinion 03/2013 on Purpose Limitation* (WP 203) at 5.

[77] GDPR, Article 5(1)(b).

[78] Merel E. Koning, 'Purpose Limitation' [2015] PI lab 4 at 6.

[79] Ibid.

[80] Ibid.

[81] GDPR, Article 6(4) read in conjunction with Recital 50.

[82] Ibid.

[83] Ibid.

purposes.[84] The change of the purpose with the data subject's consent, which must fulfil Article 7 GDPR, does not interfere with the principle of purpose limitation, as the data subject is effectively broadening the original purposes.[85] Indeed, further processing on the basis of consent means that the data subject is "in control and is responsible for determining whether the processing activity envisaged is appropriate and desirable with regards to their interests, rights, and freedoms."[86]

Recital 50 states that when the processing is compatible with the original purposes, "no legal basis separate from that which allowed the collection of the personal data is required." The literal interpretation of the provision shows that the legislator did not intend to add the requirement for a legal basis when the purposes of the processing activities are deemed compatible with the original purpose. A number of scholars ascertain that the principle of lawfulness is not undermined, as Article 6(4) specifically sets out the conditions to ensure lawfulness of further processing.[87] Nonetheless, the literal interpretation of Recital 50 is heavily contested, with commentators arguing that in addition to the fulfilment of Article 6(4), a legal basis separate from that which allowed the initial collection of the personal data is required.[88] It follows that these substantial discrepancies in understanding the necessary steps to be fulfilled

[84] GDPR, Article 5(1)(b) read in conjunction with Recital 50.

[85] Also acknowledged in Lukas Feiler, Nikolaus Forgó and Michaela Weigl, *The EU General Data Protection Regulation (GDPR): A Commentary* (German Law Publishers 2017) at 86.

[86] Wouter Seinen, Andre Walter and Sari van Grondelle, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' in Manel Medina, Andreas Mitrakas and others (eds), *Privacy Technology and Policy* (Springer 2018) at 156.

[87] See, for example: Philipp Krame, 'Art. 6 Rechtmäßigkeit der Verarbeitung' in Martin Eßer, Philipp Kramer and Kai von Lewinski (eds), *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz und Nebengesetze* (Heymanns 2018) at 59; Wouter Seinen, Andre Walter and Sari van Grondelle, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' in Manel Medina, Andreas Mitrakas and others (eds), *Privacy Technology and Policy* (Springer 2018) at 157; Tim Wybitul, 'Erlaubnistatbestände der DSGVO' in Tim Wybitul, *EU-Datenschutz-Grundverorordnung* (Fachmedien Recht und Wirtschaft 2017).

[88] See, for example: Horst Heberlein, 'Art. 6 Rechtmäßigkeit der Verarbeitung'in Eugen Ehmann and Martin Selmayr, *Datenschutz-Grundverordnung* (LexisNexis 2018) at 48; Jan Philipp Albrecht, 'Artikel 6' in Spiros Simitis, Gerrit Hornung and Indra Spiecker (eds), *Datenschutzrecht DSGVO mit BDSG* (Nomos 2019) at 13; Benedikt Buchner and Thomas Petri, 'Art. 6 Rechtmäßigkeit der Verarbeitung' in Jürgen Kühling and Benedikt Buchner (eds), Datenschutz-Grundverordnung/BDSG (Beck 2018) at 183; Philipp Reimer, 'Artikel 6 Rechtmäßigkeit der Verarbeitung' in Gernot Sydow (ed), *Europäische Datenschutzgrund-verordnung* (Nomos 2018) at 67.

to comply with the principle of lawfulness in case of further processing render the application of Article 6(4) uncertain.

4.2.1 Compatibility of Crawling Purposes

If the primary data processing activity fulfils the requirements of lawfulness, a controller intending to further process the previously collected data should carry out a compatibility assessment of the new purposes in relation to the original ones.[89] This means that, when a controller intends to crawl a specific social media platform, it must evaluate whether the purpose of the crawling is compatible with the initial purposes of the processing by the social media platform. The assessment of compatibility lies on the assumption that the purposes of the primary processing are in line with the GDPR, especially as regards specificity of the principle of purpose limitation.

In order to establish whether the new purpose in compatible with the original one the following factors can be taken into account:

i. Any link between the purposes for which the personal data have been collected and the purposes of the further processing intended;[90]
ii. The context in which the personal data have been collected, in particular the reasonable expectations of data subjects, based on their relationship with a data controller, as to their further processing;[91]
iii. The nature of the personal data, in particular whether special categories of personal data or personal data related to criminal convictions and offenses are processed;[92]
iv. The possible consequences of the intended further processing for data subjects;[93] and
v. The existence of appropriate safeguards, which may include encryption or pseudonymisation.[94]

[89] Ibid.

[90] GDPR, Article 6(4)(a).

[91] GDPR, Article 6(4)(b).

[92] GDPR, Article 6(4)(c).

[93] GDPR, Article 6(4)(d).

[94] GDPR, Article 6(4)(e).

When discussing social media crawling, it may be useful to assess compatibility on the basis of an example. It is common for a business to use crawling techniques to monitor social media sites as an insight tool to collect information on how a service, or a product, is perceived by customers. For example, a business may be interested in crawling Facebook to check customers' reviews or check whether their services are mentioned by users of the site. In order to carry out such a further processing activity, the business, acting as the controller, must assess whether the purpose of the crawling is compatible with the original purposes in Facebook's data policy. In its Data Policy, Facebook includes as a purpose that of providing measurements, analytics and other business services.[95] According to this purpose, Facebook uses the information to "help advertisers and other partners measure the effectiveness (…) of their services, and understand the types of people who use their services and how people interact with their websites, apps and services."[96] However, it must be emphasised that such a purpose may not be specific enough to fulfil the principle of purpose limitation as it should not be "vague or general, such as for instance 'improving users' experience' (…)."[97] It follows that in applying the requirements of Article 6(4) GDPR, it can be concluded that there is a link between the purposes as they similarly refer to the effectiveness of a service. It is crucial to determine whether the crawling falls within the reasonable expectations of the data subject. This should be assessed on a case-by-case basis, but the secondary controller may argue it is within the data subject's expectation that their personal data will be processed for such marketing reasons, after the subject has consented to Facebook's Data Policy. As such, the controller may also argue that the crawling itself does not have negative consequences for data subjects. Finally, it may be easier to prove compatibility of crawling if appropriate safeguards are put in place, such as pseudonymisation of personal data. Therefore, in principle, the example of monitoring of customers' review may be compatible with Facebook's original purposes. Nevertheless, the qualitative assessment of compatibility shall be carried out by balancing these five factors in relation to the specific characteristics of the further processing activity. [98] In general, when purposes of further processing are very different from the original one, they would fall beyond the reasonable expectations of the data subject and, hence, be incompatible.

---

[95] Facebook, Data Policy (*Facebook*, 2019) <https://www.facebook.com/policy.php> accessed 12 December 2019.

[96] Ibid.

[97] Article 29 Working Party, *Opinion 03/2013 on Purpose Limitation* (WP 203, 2013) at 16.

[98] Wouter Seinen, Andre Walter and Sari van Grondelle, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' [2018] at 155.

4.2.2 Crawling for Research (Purposes)

The significance of personal data for scientific research, and the natural tensions between data protection law and the needs of the research community have been reflected in the scholarship.[99] It is however worth expanding this field of scholarship to the specific field of social media scraping.

The GDPR in its expression of the purpose limitation principle (Article 5(1)(b)) provides a clause on the compatibility of purposes, as long as the new, secondary purpose falls within the scope of archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, and the processing is line with Article 89(1) GDPR. Article 89(1) GDPR requires additional safeguards to be applied in order to secure rights and freedoms of the data subject.[100] Special attention is given to the principle of data minimisation and the necessity to pseudonymise data whenever possible.[101]

The addressees of those provisions are not strictly defined in the GDPR. It is however worth mentioning, that the GDPR does not refer to researchers, or to entities engaged in research, but to research purposes. This suggests that the scope of the research privilege is not delineated by the subjects qualifying as researchers/research entities, but by activities that qualify as having a scientific or historical research purpose. Additionally, it should be noted that, as argued by some of the scholarship, Recital 159 GDPR implies "a broad understanding of the term

[99] Miranda Mourby and others, 'Are "Pseudonymised" Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK' [2018] 34 Computer Law & Security Review 222; Marc Cornock, 'General Data Protection Regulation (GDPR) and Implications for Research' [2018] 111 Maturitas A1; Kärt Pormeister, 'Genetic Data and the Research Exemption: Is the GDPR Going Too Far?' [2017] 7 International Data Privacy Law 137; Miranda Mourby and others, 'Governance of Academic Research Data under the GDPR—Lessons from the UK' [2019] 9 International Data Privacy Law 192; John Mark Michael Rumbold and Barbara Pierscionek, 'The Effect of the General Data Protection Regulation on Medical Research' [2017] 19 Journal of Medical Internet Research; Ciara Staunton, Santa Slokenberga and Deborah Mascalzoni, 'The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks' [2019] 27 European Journal of Human Genetics 1159; Gauthier Chassang, 'The Impact of the EU General Data Protection Regulation on Scientific Research' [2017] 11 Ecancermedicalscience.

[100] Gauthier Chassang, 'The Impact of the EU General Data Protection Regulation on Scientific Research' [2017] 11 Ecancermedicalscience at 8.

[101] ibid.

'research'."[102] It covers various fields of scholarship, regardless of the source of funding (i.e. public vs private).[103] As argued by some scholars, it should also include commercial research.[104]

However, it remains unclear as to what the exact scope of the scientific or historical research purpose (hereinafter: research purpose) should be. The GDPR itself does not specify the content of the term 'research purposes', making it necessary to apply standards originating from a different source. The Frascati Manual identifies five criteria that qualify a given activity as research: it must be novel, creative, uncertain, systematic, transferable and/or reproducible.[105] The EDPS, for the purposes of one of its preliminary opinions, defines scientific research as an activity where: (i) personal data are processed, (ii) relevant sectoral standards of methodology and ethics apply, including the notion of informed consent, accountability and oversight and (iii) the research is carried out with the aim of growing society's collective knowledge and wellbeing, as opposed to serving primarily one or several private interests.[106] The last requirement, is derived from AG's opinion in one CJEU case.[107] It can be further extended upon

---

[102] Konrad Lachmayer and Eva Souhrada-Kirchmayer, 'Datenschutzrecht in der wissenschaftlichen Forschung' [2018] 5 Zeitschrift für Hochschulrecht, Hochschulmanagement und Hochschulpolitik 153 at 154; Lothar Gamper / Markus Kastelitz, ,'Auswirkungen der Datenschutz-Grundverordnung auf die Wissenschaftliche Forschung in Österreich' [2018] Jusletter IT; Sebastian J Golla, Henning Hofmann and Matthias Bäcker, 'Connecting the Dots' (2018) 42 Datenschutz und Datensicherheit - DuD 89 at 90.

[103] GDPR, Recital 159.

[104] Kärt Pormeister, 'Genetic Data and the Research Exemption: Is the GDPR Going Too Far?' [2017] 7 International Data Privacy Law 137 at 140; Sebastian J Golla, Henning Hofmann and Matthias Bäcker, 'Connecting the Dots' (2018) 42 Datenschutz und Datensicherheit - DuD 89 at 90; Benedikt Buchner and Marie-Theres Tinnefeld, 'Art. 89 DS-GVO' in Jürgen Kühling and Benedikt Buchner (eds), *Datenschutz-Grundverordnung / BDSG. Kommentar* (2nd edn, 2018) at 9.

[105] *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities* (OECD Publishing, 2015) at 45. The Frascati Manual, the international standard for the use of research and experimental development statistics (R&D), requires the satisfaction of five core criteria: novel, creative, uncertain, systematic, transferable and/or reproducible.

[106] European Data Protection Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research* (2020) at 12.

[107] Opinion of AG Mancini in Case 234/83 *Gesamthochschule Duisburg v Hauptzollamt München-Mitte* [1985] on interpretation of 'scientific activities' in the context of the legislation relating to custom duties (first indent of Article 3(2) of Regulation No 1798/75) states that: 'scientific activities must be interpreted as including activities carried on by a public or private establishment engaged in education or research for the purpose of further the acquisition, development, exposition or dissemination of scientific knowledge (…)'.

by the argument present in the scholarship that research needs to be distinguished from simple data analysis.[108]

The nature of the norm introduced in the second half of the sentence in Article 5(1)(b) GDPR is unclear. One possible interpretation is that this norm establishes an automatic "mechanism": once personal data is further processed for inter alia research purposes it shall always be deemed compatible with the original purpose for which data was collected. This is in line with the literal reading of Article 5(1)(b) GDPR. However, it could also lead to the risk of controllers circumventing the purpose limitation principle by claiming that their further processing is carried out for, for example, research purposes.

It has been also argued that Article 5(1)(b) GDPR introduces a "so-called presumption of compatibility", which seems to suggest the possibility of a rebuttal.[109] Following this line of reasoning, it is yet unclear under which criteria should the presumption (potentially) be challenged or defended. Trying to accommodate the Frascati or EDPS criteria as the baseline for a possible rebuttal would lead to conceptual and practical challenges. First, going against at least the Frascati criteria in designing research would lead to an inability to produce useful information. It is therefore hardly arguable that entities would engage in research ignoring appropriate methodology as this would render their activity pointless.

However, the ethical component of the research definition as introduced by the EDPS, can have significant consequences if broadly acknowledged. Requiring that only ethically viable research activities could enjoy the research privileges the GDPR has to offer, would lead to practical challenges of enforcement. Especially in the context of private entities engaged in R&D this would require further (public) debate about in how far should the supervisory authorities or civil courts be allowed to force private entities to disclose their business secrets. This debate resembles the ongoing debate on the explainability of automated decision-making systems, where the business secrets, or IP argument, is often present. Furthermore, this would open up to additional scrutiny towards the research-engaged entities in the EU as it would require the

---

[108] Stefan Knotzer, *Wissenschaftliche Forschung und Datenschutz: Eine kritische Analyse Ausgewählter Aspekte der Österreichischen Rechtslage* (ZTR 2018) at 203.

[109] European Data Protection Board, *Opinion 3/2019 Concerning the Questions and Answers on the Interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (Art. 70.1.b)* (EDPB Guidelines 2019) at 8.

supervisory authorities (established to enforce data protection law) to assess whether a given research conduct meets the ethical (not legal) standards.

Concluding, the exact legal character of the second half of the sentence in Article 5(1)(b) GDPR remains unclear for now. Further clarifications from the courts or supervisory authorities are to be expected. Until then, it needs to be pointed out that the term research purposes is to be interpreted broadly, as to include commercial research. It is also crucial to apply the widely acknowledged criteria to define what constitutes research and what does not, such as the Frascati criteria. It follows that the question of the enforcement of compatibility of research purposes remains uncertain.

4.2.3 Incompatibility of Crawling Purposes

Under Article 6(4) GDPR, if the new purposes of the further processing are incompatible, the controller can carry out the processing only if it does so on the basis of the subject's consent or of Union or Member State law with specific safeguards. As can be ascertained from the legislative history of the provision, the intention of the legislator was to limit the legal basis for further processing to these two grounds. In other words, it is not possible for a secondary controller to rely on "broader" grounds such as legitimate interests.[110] One may argue that the threshold for further processing when the new purpose is incompatible is higher, as it is restricted to consent and specific Union or Member State law. It follows that it limits the scenario where there is a "chain of further processing activities" which are incompatible with the original purpose without the involvement of the data subject. In particular, as mentioned previously, consenting to the further processing activity expands the original purposes. Hence, it ensures that the reasonable expectations of data subjects are respected by giving them a voice in accepting or rejecting the new purposes. If the controller fulfils either of the two grounds in Article 6(4) GDPR and the processing includes special category of personal data, then, similarly to the previous discussion, compliance with Article 9(2) GDPR must be assured.

Conclusion

[110] As embedded in Article 6(1)(f) GDPR.

Social media is made up of different categories of personal data and, thus, is a vast source of information for third parties, whether for commercial, research or other further processing purposes. Given the amount, detail and possible sensitivity of the information available on social media platforms, the appropriate legal basis for social media scraping activities needs to be confirmed.

As highlighted in this chapter, the GDPR recognises the legislative distinction between processing for statistical, scientific and historic research purposes and other purposes. The GDPR also sets out different approaches for obtaining the correct legal basis pursuant to the purpose(s) and whether primary or further processing is taking place. The general initial processing of social media data may be based on the consent of the data subject or the legitimate interests of the controller. When the processing of special categories of personal data occurs pursuant to Article 9(1) GDPR, then a legal basis in Article 6(1) and 9(2) GDPR must both be demonstrable. In this case, Article 9(2) GDPR complements Article 6(1) GDPR by providing additional safeguards which are necessarily reflective of the type of data as a 'special category'.

The question of the lawful basis of social media data as further processing is a matter of whether one of the three requirements is fulfilled. Namely if the controller obtains consent for further processing from the data subject, there is a specific legal grounds in EU or national law, or the new purpose is deemed compatible with the original purposes. In assessing compatibility, a test must be performed which considers the link between the original purposes and the secondary purpose, the context of processing, the type and nature of the data, the possible consequences of the further processing, and the presence of the appropriate safeguards for processing. However, if processing is for the purpose of statistics, or scientific or historical research, then "it is not necessary to run the compatibility test"[111] and no additional legal basis for processing needs to be sought.

As discussed in this chapter, the lawfulness of social media scraping activities is closely linked to the interpretation of the principle of purpose limitation. Indeed, a controller must determine whether the processing of social media personal data fulfils Article 6(4) GDPR, as further processing. However, among data protection scholars, relevant discrepancies concerning the

[111] 'Can we use data for another purpose?' (*European Commission*) <https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/purpose-data-processing/can-we-use-data-another-purpose_en> accessed 12 February 2020.

interpretation of this provision and the principle embedded in Article 5(1)(b) GDPR render its application uncertain. It follows that there is a need to shed light on, for example, the specific definition of ‘further processing’, whether a legal basis in Article 6(1) GDPR is needed after fulfilling Article 6(4), as well as concerning the refutability of the presumption of compatibility for research purposes.

Bibliography

Primary Sources

Directive (EU) 2016/680 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 by Competent Authorities for the Purposes of the Prevention, Investigation, Detection or Prosecution of Criminal Offences or the Execution of Criminal Penalties, and on the Free Movement of such Data, and Repealing Council Framework Decision 2008/977/JHA.

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).

Treaty on the Functioning of the European Union

Secondary Sources

Article 29 Data Protection Working Party, *Advice Paper on Special Categories of Data ("Sensitive Data")*

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* (WP 217)

Article 29 Working Party, *Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679* (WP 251)

Article 29 Working Party, *Guidelines on Consent under Regulation 2016/679* (WP259, 2017)

Article 29 Working Party, *Opinion 03/2013 on Purpose Limitation* (WP203 2013)

Bart W. Schermer, Bart Custers and Simone van der Hof, ‘The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection’ [2014] 16 Ethics and Information Technology

Benedikt Buchner and Marie-Theres Tinnefeld, ‘Art. 89 DS-GVO’ in Jürgen Kühling and Benedikt Buchner (eds), *Datenschutz-Grundverordnung / BDSG. Kommentar* (2nd edn, 2018)

Benjamin Bergemann, ‘The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection’ in Marit Hansen, Eleni Kosta, Igor Nai-Fovino and Simone Fischer-Hübner (eds), *Privacy and Identity Management. The Smart Revolution* (Springer 2018)

Benjamin Bremert, ‘Legal Aspects of Text Mining Publicly Available Data’ [2017] ULD, Independent Centre for Privacy Protection

Bogdan Batrinca and Philip C. Treleaven, 'Social Media Analytics: a Survey of Techniques, Tools and Platforms' [2015] 30(1) AI & Society 89-116

Cambridge Dictionary, 'Purpose' (*Cambridge Dictionary*) <https://dictionary.cambridge.org/dictionary/english/purpose> accessed 11 December 2019

'Can we use data for another purpose?' (*European Commission*) <https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/purpose-data-processing/can-we-use-data-another-purpose_en> accessed 12 February 2020

Casey Fiesler and Nicholas Proferes, '"Participant" Perceptions of Twitter Research Ethics' [2018] Social Media + Society 1-14

Chih-Hsing Ho, ‘Challenges of the EU General Data Protection Regulation for Biobanking and Scientific Research’ [2015] 25 Journal of Law, Information and Science 1, 84-103

Ciara Staunton, Santa Slokenberga and Deborah Mascalzoni, ‘The GDPR and the Research Exemption: Considerations on the Necessary Safeguards for Research Biobanks’ [2019] 27 European Journal of Human Genetics 1159

David Kampert, ‘Artikel 9 Verarbeitung Besonderer Kategorien Personenbezogener’ Daten in Gernot Sydow (ed), *Europäische Datenschutzgrund-verordnung* (Nomos 2018)

Edward S. Dove, ‘The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era’ [2018] 46 The Journal of Law, Medicine & Ethics 4, 1013-1030

Eike Michael Frenzel, ‘Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten’ in Boris Paal and Daniel Pauly (eds*), Datenschutz-Grundverordnung Bundesdatenschutzgesetz* (Beck 2018)

Elena Gil González and Paul de Hert, ‘Understanding the Legal Provisions that Allow Processing and Profiling of Personal Data – An Analysis of GDPR Provisions and Principles’ [2019] 19 ERA Forum 4

Eleni Kosta, ‘Consent in European Data Protection Law’ [2013] Martinus Nijhoff Publishers

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* (EDPB Guidelines, 2019)

European Data Protection Board, *Guidelines 3/2019 on Processing of Personal Data through Video Devices* v2.0 (EDPB Guidelines, 2020)

European Data Protection Board, *Opinion 3/2019 Concerning the Questions and Answers on the Interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) (Art. 70.1.b)* (EDPB Guidelines 2019)

European Data Protection Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research* (2020)

Facebook, Data Policy (*Facebook*, 2019) <https://www.facebook.com/about/privacy/legal_bases> accessed 12 December 2019

Facebook, Data Policy (*Facebook*, 2019) <https://www.facebook.com/policy.php> accessed 12 December 2019

Facebook, 'What is the Face Recognition Setting on Facebook and How Does it Work?' (*Facebook*) <https://www.facebook.com/help/122175507864081> accessed 14 February 2020

Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities (OECD Publishing, 2015)

Gaby Hinsliff, 'Trash Talk: How Twitter is Shaping the New Politics' (*The Guardian*, 31 July 2016) <https://www.theguardian.com/technology/2016/jul/31/trash-talk-how-twitter-is-shaping-the-new-politics> accessed 11 December 2019

Gauthier Chassang, 'The Impact of the EU General Data Protection Regulation on Scientific Research' [2017] 11 Ecancermedicalscience

Glez-Pena and others, 'Web scraping Technologies in an API World' [2014] 15(5) Briefings in Bioinformatics 788-797

Google, 'Analytics' (*Google*) <https://analytics.google.com/analytics/web/provision/#/provision> accessed 24 February 2020

Holger Greve, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Martin Eßer, Philipp Kramer and Kai von Lewinski (eds), *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz und Nebengesetze* (Heymanns 2018)

Horst Heberlein, 'Art. 6 Rechtmäßigkeit der Verarbeitung'in Eugen Ehmann and Martin Selmayr, *Datenschutz-Grundverordnung* (LexisNexis 2018)

Jan Philipp Albrecht, ‘Artikel 6’ in Spiros Simitis, Gerrit Hornung and Indra Spiecker (eds), *Datenschutzrecht DSGVO mit BDSG* (Nomos 2019) at 13; Benedikt Buchner and Thomas Petri, ‘Art. 6 Rechtmäßigkeit der Verarbeitung’ in Jürgen Kühling and Benedikt Buchner (eds), Datenschutz-Grundverordnung/BDSG (Beck 2018)

John Mark Michael Rumbold and Barbara Pierscionek, ‘The Effect of the General Data Protection Regulation on Medical Research’ [2017] 19 Journal of Medical Internet Research

Kärt Pormeister, ‘Genetic Data and the Research Exemption: Is the GDPR Going Too Far?’ [2017] 7 International Data Privacy Law 137

Konrad Lachmayer and Eva Souhrada-Kirchmayer, ‘Datenschutzrecht in der wissenschaftlichen Forschung’ [2018] 5 Zeitschrift für Hochschulrecht, Hochschulmanagement und Hochschulpolitik 153

Leximancer, 'Text in Insight Out' (*Leximancer*) <https://info.leximancer.com/> accessed 11 December 2019

Lothar Gamper / Markus Kastelitz, ‘Auswirkungen der Datenschutz-Grundverordnung auf die Wissenschaftliche Forschung in Österreich‘ [2018] Jusletter IT; Sebastian J Golla, Henning Hofmann and Matthias Bäcker, ‘Connecting the Dots’ (2018) 42 Datenschutz und Datensicherheit - DuD 89

Lukas Feiler, Nikolaus Forgó and Michaela Weigl, *The EU General Data Protection Regulation (GDPR): A Commentary* (German Law Publishers 2017)

M. Saniya Parvez and others, ‘Analysis of Different Web Data Extraction Techniques’ [2018] International Conference on Smart City and Engineering Technology

Marc Cornock, ‘General Data Protection Regulation (GDPR) and Implications for Research’ [2018] 111 Maturitas A1

Merel E. Koning, ‘Purpose Limitation’ [2015] PI lab 4

Michael Meng and others, 'Application Programming Interface Documentation: What Do Software Developers Want?' [2018] 48(3) Journal of Technical Writing and Communication 295-330

Miranda Mourby and others, ‘Are “Pseudonymised” Data Always Personal Data? Implications of the GDPR for Administrative Data Research in the UK’ [2018] 34 Computer Law & Security Review 222

Peter Schantz, ‚Zulässigkeit der Verarbeitung Besonderer Kategorien Personenbezogener Daten‘ in Peter Schantz and Heinrich Amadeus Wolff (eds), *Das Neue Datenschutzrecht* (Beck 2017)

Philipp Krame, ‘Art. 6 Rechtmäßigkeit der Verarbeitung’ in Martin Eßer, Philipp Kramer and Kai von Lewinski (eds), *Datenschutz-Grundverordnung, Bundesdatenschutzgesetz und Nebengesetze* (Heymanns 2018)

Philipp Reimer, ‘Artikel 6 Rechtmäßigkeit der Verarbeitung’ in Gernot Sydow (ed), *Europäische Datenschutzgrund-verordnung* (Nomos 2018)

Ray Poynter, *The Handbook of Online and Social Media Research: Tools and Techniques for Market Researchers* (1st edn, John Wiley & Sons 2012) 223

Sebastian J Golla, Henning Hofmann and Matthias Bäcker, ‘Connecting the Dots’ (2018) 42 Datenschutz und Datensicherheit - DuD 89

Sebastian Schulz, ‘Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten’ in Peter Gola (ed), *Datenschutz-Grundverordnung* (Beck 2018)

Spyros Polykalas, ‘Assessing General Data Protection Regulation for Personal Data: is the End of “Take it or Leave it” Approach for Dowloading Apps?’ [2017] The Seventh International Conference on Social Media Technologies, Communication, and Informatics

Stefan Knotzer, *Wissenschaftliche Forschung und Datenschutz: Eine kritische Analyse Ausgewählter Aspekte der Österreichischen Rechtslage* (ZTR 2018)

Thilo Weichert, 'Art. 9 Verarbeitung Besonderer Kategorien Personenbezogener Daten' in Jürgen Kühling and Benedikt Buchner (eds), Datenschutz-Grundverordnung/BDSG (Beck 2018)

Thomas Petri, 'Artikel 9' in Spiros Simitis, Gerrit Hornung and Indra Spiecker (eds), *Datenschutzrecht DSGVO mit BDSG* (Nomos 2019)

Tim Adams, 'Facebook's Week of Shame: the Cambridge Analytica Fallout' (*The Guardian*, 24 March 2018) <https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence> accessed 11 December 2019

Tim Wybitul, 'Erlaubnistatbestände der DSGVO' in Tim Wybitul, *EU-Datenschutz-Grundverorordnung* (Fachmedien Recht und Wirtschaft 2017)

Twitter, 'Advanced Search' (*Twitter*) <https://twitter.com/search-advanced?lang=en> accessed 11 December 2019

Twitter, 'Privacy Policy' (*Twitter*, 2019) <https://twitter.com/en/privacy> accessed 12 December 2019

Twitscoop, 'Worldwide Twitter Trends' (*Twitscoop*) <https://www.twitscoop.com/> accessed 24 February 2020

Wouter Seinen, Andre Walter and Sari van Grondelle, 'Compatibility as a Mechanism for Responsible Further Processing of Personal Data' in Manel Medina, Andreas Mitrakas and others (eds), *Privacy Technology and Policy* (Springer 2018)

Yaniv Taigman and others, 'DeepFace: Closing the Gap to Human-Level Performance in Face Verification' [2014] Facebook Research, Conference on Computer Vision and Pattern Recognition (CVPR)