Computer Law & Security Review 52 (2024) 105956

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

Tell me something new: data subject rights applied to inferred data and profiles

Bart Custers[a,*], Helena Vrabec[a,b]

[a] eLaw, The Center for Law and Digital Technologies at The Leiden Law School, Leiden University, The Netherlands
[b] Maplebear Inc, d/b/a Instacart, San Francisco, USA

ARTICLE INFO

Keywords:
Data subject rights
GDPR
Data protection
Inferred data
Transparency
Profiling
Automated decisions
Automated decision-making

ABSTRACT

The EU General Data Protection Regulation (GDPR) contains several data subject rights, but for many of these rights it is not entirely clear how they should work in practice, especially in digital environments. Most data subject rights apply to personal data obtained directly or indirectly from the data subject. This is often personal data that data subjects already are familiar with, i.e., things they already know about themselves. Unclear, however, is to what extent ascribed personal data, such as inferred data and categories or profiles in which data subjects are placed by data controllers, are within the scope of these rights. Such ascribed personal data often concerns novel information, generated by data controllers, and includes insights into how controllers view and assess them, which may have practical and legal impact on data subjects. Given these characteristics, the ascribed personal data may be much more interesting to data subjects, so it appears beneficial, from the policy perspective, to have this novel information included in the scope of data subject rights. If data subject rights do not apply to inferred data and profiles, invoking these rights is unlikely to be informative and provide meaningful information for data subjects, particularly in complex, digital environments. However, if data subject rights do apply to inferred data and profiles, the scope of these rights may be hard to delineate and they may quickly interfere with rights and freedoms of others, including trade secrets of data controllers and privacy rights of other data subjects. In this article, we investigate the implications of applying data subject rights to inferred data and profiles. For each data subject right in the GDPR, we assess which types of personal data could and perhaps should be in scope, based on grammatical and teleological legal analyses as well as practical considerations. While the area of data subject rights received significant academic attention in the past years, our article contributes to the discussion by providing a systematic, holistic framework to consider the scope of the rights in relation to ascribed data.

1. Introduction

In the data economy, large amounts of personal data are collected and processed.[1] This includes data that people provide themselves to companies and governments and data that are collected on them without being aware of this.[2] Apart from collecting existing data, also new data on people is generated through data analytics, often by means of inferring characteristics from available data, creating categories in which people are placed, or creating (risk) profiles to which people are assigned.

The EU General Data Protection Regulation (GDPR)[3] contains several data subject rights that people can invoke, for instance, if they want to

* Corresponding author.
E-mail address: b.h.m.custers@law.leidenuniv.nl (B. Custers).

[1] See, for instance, Elisabetta Raguseo, Big data technologies: An empirical investigation on their adoption, benefits and risks for companies, International Journal of Information Management, Volume 38, Issue 1, 2018, Pages 187-195.

[2] In 2023, Google was sued for tracking users' activity when they set Google's Chrome browser to "Incognito" mode and other browsers to "private" browsing mode. Jonathan Stempel,Google settles $5 billion consumer privacy lawsuit, December 29, 2023, https://www.reuters.com/legal/google-settles-5-billion-consumer-privacy-lawsuit-2023-12-28/. In the aftermath of the Dobbs decision by the US Supreme Court, concerns have been raised about the law enforcement access to the data held by pharmacy chains that have access to a person's medical records across multiple states and could be pulled by prosecutors seeking to clamp down on people who help women travel.

[3] Regulation (EU) 2016/679 (General Data Protection Regulation) in the current version of the OJ L 119, 04.05.2016.

https://doi.org/10.1016/j.clsr.2024.105956

Available online 18 March 2024

Table 1
To which categories of personal data do data subject rights apply.

Data subject right	Directly obtained data	Indirectly obtained data	Inferred data	Categories and profiles
Right to information (Art. 12-14)	Yes (Art. 13)	Yes (Art. 14)	Ambiguous (possible under Art. 14)	Partially (Art. 13.2.f and 14.2.g)
Right of access (Art. 15)	Yes	Yes	Partially (Recital 63)	Partially (Art. 15.1.h)
Right to rectification (Art. 16)	Yes	Yes	No	No
Right to erasure (Art. 17)	Yes	Yes	Yes	No
Right to restriction of processing (Art. 18)	Yes	Yes	Yes (only through the reference in Art. 18.1.d)	Yes (only through the reference in Art. 18.1.d)
Right to data portability (Art. 20)	Yes	Yes	Likely no	No
Right to object/ automated decisions (Art. 21-22)	Yes	Yes	Yes	Yes

object to particular practices. However, for many of these rights it is not clear how they should work in practice. Most data subject rights seem to apply to *all* personal data obtained directly or indirectly from the data subject. But in practice it is often unclear to what extent ascribed personal data, such as inferred data and categories or profiles in which data subjects are placed by data controllers, are also within the scope of these rights.[4]

Since the scope of data subject rights is limited to personal data, the obvious starting point would be to assess whether these newly generated data are indeed personal data. If ascribed to individuals, they usually are personal data.[5] If they are only generic models, such as group profiles (e.g., people who buy beer are also likely to buy diapers, or people driving white cars are more prone to attract colon cancer), it may be hard to determine at which point these characteristics apply to individual data subjects and can be considered personal data.[6] Typically, a profile like 55 % of all teenagers are gamers is not personal data (no identified or identified person is mentioned), but becomes personal data when ascribed to John, who is a teenager (the probability that John, who is a teenager, is a gamer is 55 %).

But even when it is clear that these inferred data and profiles are personal data, it depends on the exact phrasing of each data subject right whether it can be applied. In other words, apart from interpreting whether (and at which point) inferred data and profiles are personal data, also each data subject right needs to be interpreted to determine whether it covers inferred data and profiles. Much depends on the interpretation of these provisions (including the provisions on the exceptions to data subject rights), which can be done in different ways. The most common interpretative methods in law are the grammatical (or textual) interpretation, the systematic (or contextual) interpretation, the teleological (or purposive) interpretation, and the historical interpretation.[7] In this article, for each data subject right in the GDPR it is assessed which types of personal data could and perhaps should be in scope, based on grammatical, contextual and teleological legal analyses and practical considerations.

The goal of these analyses is to find the right balance between narrow and broad interpretations of data subject rights. Narrow interpretations,[8] often focusing on grammatical interpretations, would only put lower order types (i.e., data that can be obtained without the use of data analytics) of personal data like (raw) data collected directly or indirectly from data subjects within the scope of data subject rights. Broad interpretations, with more focus on teleological interpretations, would also put higher order types (i.e., data that can only be obtained using sophisticated data analytics) of personal data like inferred data, predictions and (risk) profiles within the scope of data subject rights. Narrow interpretations, which businesses appear to favor,[9] may be unlikely to yield informative and meaningful information for data subjects. In other words, it may not tell them anything they did not already know. Broad(er) interpretations may be more insightful and meaningful for data subjects and may be more in line with the set-up and goals of the GDPR. However, when inferred data and profiles are also within the scope of data subject rights, this may quickly interfere with rights and freedoms of others, including trade secrets of data controllers, the freedom of doing business and privacy rights of other individuals.[10] It is exactly the knowledge resulting from (sometimes sophisticated) data analytics that contains added value for data controllers. Typically, the tools for data analytics (including algorithms) and the knowledge resulting from it (such as profiles) constitute the competitive edge of companies. These companies may suggest that profiles and other knowledge distilled from personal data are corporate secrets because they may, *via* reverse engineering, enable disclosure of their analyses and software.[11] Therefore, broad(er) interpretations may give rise to complications and may be much harder to delineate.[12] In order to take a more qualified and detailed approach to this, each data subject right is analyzed separately in this article.

The area of data subject rights has received significant attention in

[4] Interestingly, the US state privacy laws (in California, Colorado and Washington, to mention a few) have been more explicit about the inclusion of inferred data in the definition of personal information. Blanke, J. M. “Protection for ‘Inferences drawn’: A comparison between the general data protection regulation and the California consumer privacy act”. In: Global Privacy Law Review 1.2 (2020), pp. 81–92. url: https: //kluwerlawonline.com/journal-article/Global+Privacy+Law+Review/1.2/GPLR2020080.

[5] Assuming the individuals are identified or identifiable.

[6] For a critical analysis of the binary approach to defining personal data see: El Khoury, Alessandro, Personal Data, Algorithms and Profiling in the EU: Overcoming the Binary Notion of Personal Data through Quantum Mechanics (December 1, 2018). Erasmus Law Review, Vol. 11, No. 3, 2018, Available at SSRN: https://ssrn.com/abstract=3356767.

[7] Ammann, O. (2020) The Interpretative Methods of International Law: What Are They, and Why Use Them? In: Ammann, O. (ed.) *Domestic Courts and the Interpretation of International Law*. Leiden, The Netherlands: Brill | Nijhoff.

[8] These are interpretations which the European Data Protection Board (EDPB) and the Court of Justice of the EU (CJEU) appear to support in some cases. EDPB, Guidelines 01/2022 on data subject rights - Right of access, 18 January 2022, page 31. Sandra Wachter and Brent Mittelstand, A right to reasonable inferences: re-thinking data protection law in the age of big data and AI, Columbia Business Law Review, Vol 2019, No, 2.

[9] This seems to be the interpretation that some tech companies have embraced. See Helena U. Vrabec (2021) Data subject rights under the GDPR, Oxford University Press, p. 214. Also see Mantelero, Alessandro, The EU Proposal for a General Data Protection Regulation and the Roots of the ‘Right to Be Forgotten’ (June 28, 2013). Computer Law & Security Review, Volume 29, Issue 3, June 2013, p.11, Available at SSRN: https://ssrn.com/abstract=2473151.

[10] For instance, in the US, the industry association Technet, has argued against the adoption of a broad access right in the Oregon state privacy law. Technet’s letter as of March 7, 2023, to the Oregon state legislature; https://olis.oregonlegislature.gov/liz/2023R1/Downloads/PublicTestimonyDocument/60649. Also see: Case C-579/21, Pankki, ECLI:EU:C:2023:501, recital 80.

[11] Hildebrandt, M. (2011) *The Rule of Law in Cyberspace?* Inaugural Lecture, Nijmegen: Radboud University. https://works.bepress.com/mireille_hildebrandt/48/.

[12] Suggesting that the broad definition of personal data is challenging to work with, Lorenzo Dalla Corte, Scoping Personal Data: Towards a Nuanced Interpretation of the Material Scope of EU Data Protection Law, The European journal of Law and Technology, Vol 10 No 1 (2019) https://ejlt.org/index.php/ejlt/issue/view/59.

literature over the past years.[13] This article intends to contribute to the discussion by providing a systematic, comprehensive, and practical holistic framework to consider the scope of data subject rights in relation to ascribed data. To achieve this, a taxonomy of different types of personal data is proposed. Subsequently, for each data subject right the scope regarding each type of personal data in the taxonomy is assessed.

This article is structured as follows. Section 2 proposes a taxonomy of different types of personal data that play a role in providing transparency for data subjects and empowering them with regard to the processing of their personal data. Section 3 examines for each data subject right in the GDPR (i.e., articles 12 through 22) to which of the types of personal data identified in Section 2 it applies, mainly by looking at the phrasing of the GDPR provisions. Section 4 analyses these findings and contrasts them with a more teleological and contextual approach, considering the intentions and structure of the GDPR. Section 5 provides conclusions.

2. Types of personal data

According to Article 4(1) of the GDPR, personal data means any information relating to an identified or identifiable natural person (the 'data subject). This definition includes data that data subjects know about themselves and are usually familiar with, such as their address, date of birth, nationality, or marital status. This data is often rather factual, even though it can change over time. Some of these data, though factual, can be more subjective, such as personal preferences, hobbies, or interests. Data like these are often collected from the data subjects directly or collected on them indirectly, *via* data brokers.[14]

The definition of personal data also includes data that data subjects do not know about themselves or are less familiar with.[15] This is when data controllers start ascribing characteristics to a data subject. Such characteristics can sometimes be inferred from existing data, such as someone's age which can be inferred easily from a date of birth. Most of these characteristics are statistical in nature, as they often contain predictions. Typically, based on available personal data, predictions can be made about someone's preferences (e.g., based on searches on Google or Amazon, it can be predicted what else you may be interested in). Also a data subject's health, credit score, or life expectancy can be predicted in such ways and ascribed to data subjects, for instance, by insurance companies or banks.[16] Data subjects may not know such information about themselves or may not be familiar with it. Typically, people do not know their (personal) life expectancy and the predictions may reveal interests or preferences that people were not previously aware of.[17] Data like these are often not collected from data subjects but generated by data controllers.

Inferring someone's age from a date of birth can be done correctly without much hassle, but obviously predictions on personal preferences or health status may have limited reliability. Some of these predictions may be very sophisticated and with high accuracy, others may be flawed and with limited accuracy.[18] It is important to keep in mind that the

[13] By way of example, and in addition to other sources referenced in this article, these are some of the key publications on the topic: Kuner et al. (ed.), The EU General Data Protection Regulation (GDPR): A Commentary (Oxford University Press 2020); Erdos, David, Comparing Constitutional Privacy and Data Protection Rights within the EU (May 11, 2021). University of Cambridge Faculty of Law Research Paper No. 21/2021, https://ssrn.com/abstract=3843653; Kamarinou, Dimitra and Millard, Christopher and Turton, Felicity, Protection of Personal Data in Clouds and Rights of Individuals (May 2021). Chapter 8, 'Protection of Personal Data in Clouds and Rights of Individuals', in C. Millard (ed.) Cloud Computing Law, (2nd edn, OUP 2021, https://ssrn.com/abstract=425583. Jef Ausloos, Michael Veale and René Mahieu, Getting Data Subject Rights Right, 10 (2019) JIPITEC 283; Li, Wenlong and Toh, Jill, Data Subject Rights as a Tool for Platform Worker Resistance: Lessons from the Uber/Ola Judgments (December 9, 2022), https://ssrn.com/abstract=4306868. -With regards to the right not to be subjected to automated decision-making: Selbst, Andrew D. and Powles, Julia, Meaningful Information and the Right to Explanation (November 27, 2017). International Data Privacy Law, vol. 7(4), 233-242 (2017), https://ssrn.com/abstract=3039125;; Noto La Diega, Guido, Against the Dehumanisation of Decision-Making – Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information (May 31, 2018). 9 (2018) JIPITEC 3 para 1, https://ssrn.com/abstract=3188080;; Malgieri, Gianclaudio, Automated Decision-Making in the EU Member States: The Right to Explanation and Other 'Suitable Safeguards' for Algorithmic Decisions in the EU National Legislations (August 17, 2018). Computer Law & Security Review, 2019 Forthcoming, https://ssrn.com/abstract=3233611; Tosoni, Luca, The Right To Object to Automated Individual Decisions: Resolving the Ambiguity of Article 22(1) of the General Data Protection Regulation (May 14, 2021). 11 International Data Privacy Law (2021) (Forthcoming), University of Oslo Faculty of Law Research Paper No. 2021-07, https://ssrn.com/abstract=3845913/.;- with regards to the right of access: Veale, Michael and Binns, Reuben and Ausloos, Jef, When Data Protection by Design and Data Subject Rights Clash (February 20, 2018). International Data Privacy Law (2018) doi:10.1093/idpl/ipy002, https://ssrn.com/abstract=3081069. -with regards to the right of information: Naudts, Laurens and Dewitte, Pierre and Ausloos, Jef, Meaningful Transparency through Data Rights: A Multidimensional Analysis (July 27, 2021). (Forthcoming), in Eleni Kosta, Ronald Leenes and Irene Kamara (eds), Research Handbook on EU data protection, Edward Elgar, 2022, Amsterdam Law School Research Paper No. 2022-37, Institute for Information Law Research Paper No. 2022-02, https://ssrn.com/abstract=3949750. ; -with regards to the right to be forgotten: Jef Ausloos. *The Right to Erasure in EU Data Protection Law*. Oxford University Press (2020), Koops, Bert-Jaap, Forgetting Footprints, Shunning Shadows: A Critical Analysis of the 'Right to Be Forgotten' in Big Data Practice (December 20, 2011). SCRIPTed, Vol. 8, No. 3, pp. 229-256, 2011, Tilburg Law School Research Paper No. 08/2012, https://doi.org/10.2139/ssrn.1986719; Tsesis, Alexander, Data Subjects' Privacy Rights: Regulation of Personal Data Retention and Erasure (January 30, 2019). University of Colorado Law Review, Vol. 90, 2019, https://ssrn.com/abstract=3325973. -with regards to the right to data portability: De Hert, Paul and Papakonstantinou, Vagelis and Malgieri, Gianclaudio and Beslay, Laurent and Sanchez, Ignacio, The Right to Data Portability in the GDPR: Towards User-Centric Interoperability of Digital Services (2018). Computer Law & Security Review (2018) 193–203, https://ssrn.com/abstract=3447060; Graef, Inge and Husovec, Martin and Purtova, Nadezhda, Data Portability and Data Control: Lessons for an Emerging Concept in EU Law (December 15, 2017). German Law Journal 2018, vol. 19 no. 6, p. 1359-1398, Tilburg Law School Research Paper No. 2017/22, TILEC Discussion Paper No. 2017-041, https://ssrn.com/abstract=3071875; Swire, Peter and Lagos, Yianni, Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique (May 31, 2013). 72 Maryland Law Review 335 (2013), Ohio State Public Law Working Paper 204, https://ssrn.com/abstract=2159157; Drechsler, Laura, Practical Challenges to the Right to Data Portability in the Collaborative Economy (June 21, 2018). Collaborative Economy: Challenges and Opportunities, Proceedings of the 14th International Conference on Internet, Law & Politics. Universitat Oberta de Catalunya, Barcelona, 21–22 June, 2018, https://ssrn.com/abstract=3296222.

[14] See for example Zuboff, Shoshana. "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization." Journal of Information Technology 30, no. 1 (March 2015): 75–89.

[15] For instance, Facebook's shadow profiles have been largely unknown to regular web users: Aguiar, Luis; Peukert, Christian; Schäfer, Maximilian; Ullrich, Hannes (2022): Facebook shadow profiles, DIW Discussion Papers, No. 1998, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin. Also see.

[16] Freya Van Den Boom, Regulating connected car data access and use, for Telematics Insurance in Europe, PhD dissertation, Bournemouth University, 2021, pp. 47-48, available at: https://eprints.bournemouth.ac.uk/36956/1/VAN%20DEN%20BOOM%2C%20Freyja_Ph.D._2021.pdf.

[17] Eric Siegel, When Does Predictive Technology Become Unethical?, Harvard Business Review, October 23, 2020, https://hbr.org/2020/10/when-does-predictive-technology-become-unethical.

[18] Custers, B.H.M. (2003) *Effects of Unreliable Group Profiling by Means of Data Mining.* In: G. Grieser, Y. Tanaka and A. Yamamoto (eds.) *Lecture Notes in Artificial Intelligence, Proceedings of the 6th International Conference on Discovery Science (DS 2003)* Sapporo, Japan. Berlin, Heidelberg, New York: Springer-Verlag, Vol. 2843, p. 290-295.

GDPR does not exclude incorrect personal data from its scope.[19] Data relating to a data subject is always personal data, regardless of whether such data is correct or incorrect. Suppose that a data controller puts in a database that John is female, even though John is by all physical, physiological, and psychological standards regarded as male by himself and others. In that case the characteristic is obviously incorrect, but since it is ascribed to John, it is still personal data and therefore within the scope of the GDPR.[20] Since the GDPR applies to this, John can invoke his right to rectification to set this straight, assuming he is aware of this data ascribed incorrectly to him.

The GDPR does not explicitly distinguish or recognize that there are different types of personal data, but implicitly it does.[21] Based on how personal data are collected and/or generated, we present different types of personal data in this section. We discuss personal data directly obtained from the data subject (Section 2.1), personal data not directly obtained from the data subject (Section 2.2), inferred data (Section 2.3), and categories and profiles in which data controllers put data subjects (Section 2.4). The former two types of personal data are data collected from or on data subjects themselves, the latter two types of personal data are data generated by data controllers.

2.1. *Personal data directly obtained from the data subject*

People generate large volumes of personal data on themselves every day, for instance, *via* the use of social media as well as by technology, including sensors (e.g., cameras, microphones), trackers (e.g., RFID tags, web surfing behavior) and other devices (e.g., mobile phones, wearables for self-surveillance/quantified self, log-in data). Such data can include data regarding location, preferences, and online behavior. Data controllers can directly obtain such data from the data subject, assuming there is a legal basis for this. The legal basis for collecting and processing personal data has to be one of the options mentioned in Article 6 of the GDPR. Most often, the legal basis will be consent of the data subject or the necessity to process the personal data for the performance of a contract between the data controller and the data subject. For instance, when using social media, users often consent to data controllers collecting and processing their personal data when they register or create an account.

Although they may have been informed about the fact that their personal data is collected and processed, and for which purposes, data subjects may not be aware of this. Privacy notices are rarely read and might not conspicuously appear at a data collection point.[22] Additionally, most of the data collecting and processing takes place beyond what people see. When people register an account or purchase products or services (either online or offline) they often have to complete several forms and provide personal details. Often the data asked for in such forms contains trivial information, such as contact details, shipping addresses, and credit card numbers. People usually consider such data innocuous and only contact data controllers when they move to a different address or when products or services are not adequately delivered. What people often fail to notice, however, is that even their minimal use of an online service has generated a vast amount of information which might have been shared with third parties such as vendors and brokers.[23]

2.2. *Personal data not directly obtained from the data subject*

Most people do not realize that data controllers have much more personal data available on data subjects than the data they obtained directly from them.[24] The current data economy is based on trading, selling, and leasing data and extracting added value from the raw data through data analysis that can reveal new information and knowledge. Large data brokers like Acxiom and Experian have built their businesses on such practices. Acxiom, for instance, claims to have files on 2.5 billion people, with about 11.000 pieces of different data attributes.[25] Obviously, most people are not aware of this and are unable to tell which pieces of information this concerns. The main reason for this is that most of these pieces of information are collected not directly from the data subject and/or yielded by data analytics (see Sections 2.3 and 2.4 below).

The GDPR explicitly acknowledges that data can be obtained either directly or indirectly from the data subject in Articles 13 and 14, which both deal with the right to information. Article 13 addresses the right to information where personal data are collected (directly) from the data subject. Article 14 addresses the right to information where personal data have not been obtained from the data subject (but have been collected indirectly). The difference is important because of practical reasons: when collecting personal data directly from a data subject, there is a direct relationship and some form of contact between the data subject and the data controller. If data is collected indirectly on a data subject, no such relationship or contact may exist. For this reason, Article 14.5.b GDPR explicitly addresses the practical complications that may follow from this by stating that the data controller does not need to provide information to the data subject if this is impossible or would involve a disproportionate effort. The distinction made in Articles 13 and 14 GDPR is understandable, but this distinction is not continued in the subsequent articles that contain other data subject rights.

2.3. *Inferred data*

In the data economy, many companies try to gain a competitive edge by extracting hidden knowledge from large amounts of data *via* data mining and machine learning.[26] This can be seen as an input-output process, in which knowledge is extracted from raw data.[27] Knowledge then refers to any information that has added value, which can be new (enriched) data at an individual level (discussed in this section) or patterns, such as profiles, at a group level (discussed in the Section 2.4).

In essence, inferred data is any new data that is inferred from already available data. The most straightforward example was already mentioned, i.e., inferring someone's age from a date of birth. From the

[19] At the most basic level, this follows from the 'a contrario' analysis of the right to correct inaccurate personal data. if inaccurate data was not personal information, then the right to rectify would not need to exist either.

[20] Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2021) A little bird told me your gender: Gender inferences in social media. *Information Processing & Management*, 58(3), 102541.

[21] For example, personal data could be categorized based on the proprietary relationship with data subjects and data users. See for instance Malgieri, Gianclaudio, Property and (Intellectual) Ownership of Consumers' Information: A New Taxonomy for Personal Data (April 20, 2016). Privacy in Germany - PinG, n. 4, 2016, 133 ff., Available at SSRN: https://ssrn.com/abstract=2916058.

[22] Woodrow Hartzog, Privacy's Blueprint, Harvard University Press, 2018, pp. 210–211.

[23] Zard L. & Sears A.M. (2023), Targeted Advertising and Consumer Protection Law in the European Union, Vanderbilt Journal of Transnational Law 56(3): p. 814.

[24] Helen Kennedy, Susan Oman, Mark Taylor, Jo Bates & Robin Steedman (2020) Public understanding and perceptions of data practices: a review of existing research. Living With Data, University of Sheffield. http://livingwithdata.org/current-research/publications/, page 6.

[25] Sherman, Justin. "Data Brokers are a Threat to Democracy," WIRED, April 13, 2021. https://www.wired.com/story/opinion-data-brokers-are-a-threat-to-democracy/.

[26] Custers, B., and Bachlechner, D. (2018) Advancing the EU Data Economy: Conditions for Realizing the Full Potential of Data Reuse, *Information Polity,* 22 (4): 291–309.

[27] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine, 17*(3), 37, 82–88. Available at: https://ojs.aaai.org/index.php/aimagazine/article/view/1230/1131.

data available on a data subject, often many other characteristics can be inferred. A typical example is the inference of personal characteristics on the basis of Facebook likes.[28] What you like (e.g., preferences regarding music, videos, people, etc.) typically reveals some personal characteristics. Music preferences are often indicative about someone's age and movie preferences can be indicators for someone's gender, simply because different generations like different types of music and men and women tend to like different types of movies. In this way, also sensitive characteristics like ethnicity, sexual orientation, religion, happiness, intelligence, and substance abuse can be predicted. These predictions may concern characteristics that people may not want to disclose about themselves (such as sexual orientation or mental health) and characteristics that people may not even know about themselves (such as the probability to attract cancer in the next decade). Such characteristics can be inferred from combinations of available data, usually not only on the data subject involved, but on data relating to many data subjects with similar characteristics. Because data-driven future forecasting is probabilistic, the inferences (i.e., new knowledge) cannot be verified after the predictions have been made.[29]

Such inferences, whether correct or not, can then be ascribed to individuals in a database, through which they become personal data. It is important to clarify the distinction between data not obtained from the data subject and inferred data. Data not obtained from the data subject are data that are indirectly obtained *from* the data subject, for instance, *via* a data broker. Inferred data are distilled from other available data (while still relating to the data subject). In other words, inferred data is newly generated data, whereas data obtained indirectly from the data subject merely involves a transfer of personal data that already existed. Typically, if a person shares her date of birth with data controller A, for this data controller this is personal data (directly) obtained from the data subject (Article 13 GDPR applies). If data controller A shares the data with data controller B,[30] for the latter this is personal data not obtained (directly) from the data subject, but indirectly, *via* data controller A. If a data controller A or B infers from the date of birth the current age of the data subject, then this age is inferred personal data. Note that if the data subject had provided her age to data controller A, it would not have been inferred data. In other words, the way in which data is obtained is decisive in this categorization, not the type of data itself. Having said that, some types of data can only be inferred and never really be provided by a data subject. Typically, personalized statistics and predictions, such as life expectancies and risks to attract particular diseases, can only be distilled from larger datasets.[31] This is not information that data subjects know about themselves.

2.4. Categories and profiles

In many situations, companies and government organizations are more interested in the characteristics of large groups of people than of specific individuals. For companies, it is interesting to identify new groups of customers, which groups of existing customers might be interested in new products and services, or which groups of customers yield the highest margins. Government organizations may be typically interested in which groups of people are in need of assistance (e.g., education, health care, social security) or are more likely to commit fraud. To identify such target groups or risk groups, organizations often use profiling techniques. This is typically done through classification or clustering.[32] Classification entails that people are placed in predefined categories, often called classes. Clustering entails that people are grouped in clusters that are not predefined. Both methods result in profiles, in which people are assigned to a specific group and the group attributes are ascribed to all group members. Ascribing group attributes to all group members is often inherently unreliable, as not all group members may share all group attributes. Particularly when group averages are ascribed to each group member, it is important to realize that each individual group member may as an individual have characteristics that deviate from the group average. Here a distinction can be made between distributive group profiles, in which each characteristic of a group profile is valid for each individual group member (even without considering the group), and non-distributive group profiles, in which a characteristic of a group profile is valid for the group and for individuals as members of that group, but not for individuals as such.[33]

Note that profiling is also a form of inferring data.[34] However, inferred data usually refers to single pieces of information, whereas profiles or categories are more like overall labels for these characteristics. A typical example is that Facebook identifies 4 types of users (relationship builders, window shoppers, town criers, and selfies).[35] All Facebook users are put into these categories, assuming they share the group characteristics and are treated accordingly. The GDPR definition of profiling (art. 4.4 GDPR) includes any form of automated processing of personal data consisting of the use of personal data to evaluate personal aspects relating to a natural person. This is a broad definition that includes both inferred data and categories and profiles. At the same time the GDPR definition of profiling does not acknowledge that profiles can also be created without the use of automated processing. Typically, profiles can also be established with traditional (i.e., non-automated) statistics or even manually (for instance, *via* observations, rules of thumb, or stereotyping).[36]

Profiling can be based on raw data, directly or indirectly obtained from the data subject, but it can also be (fully or partially) based on inferred data. Profiling based on datasets from data brokers that contain large amounts of inferred data, may propagate any existing biased patterns.[37] For instance, for profiling insurance premiums, a dataset with income data (directly obtained from data subjects) is less valuable than a dataset further enriched by the data broker with credit scores (inferred data). However, the credit scores may already be based on the income

[28] Kosinski, M., Stillwell, D., and Thore Graepel, T. (2012) Private Traits and Attributes are Predictable from Digital Records of Human Behaviour. *Proceedings of the National Academy of Sciences,* USA 110: 5802–5.

[29] Matsumi, Hideyuki and Solove, Daniel J., The Prediction Society: Algorithms and the Problems of Forecasting the Future (July 30, 2023). GWU Legal Studies Research Paper No. 2023-58, GWU Law School Public Law Research Paper No. 2023-58, Available at SSRN: https://ssrn.com/abstract=4453869, p. 6.

[30] This is allowed, for instance, if the purposes are compatible or if the data subject has consented. Ursic, H. and Custers, B.H.M. (2016) Legal barriers and enablers to big data reuse - a critical assessment of the challenges for the EU law, *European Data Protection Law Review*, Vol. 2, No. 2, p. 209–221.

[31] Matsumi, Hideyuki and Solove, Daniel J., The Prediction Society: Algorithms and the Problems of Forecasting the Future (July 30, 2023). GWU Legal Studies Research Paper No. 2023-58, GWU Law School Public Law Research Paper No. 2023-58, Available at SSRN: https://ssrn.com/abstract=4453869, p.22.

[32] Calders T. & Custers B.H.M. (2013) What is data mining and how does it work? In: Custers B.H.M., Calders T., Schermer B., Zarsky T. (eds.) *Discrimination and Privacy in the Information Society.* nr. 3 Heidelberg: Springer.

[33] Vedder, A.H. (1996) Privacy en woorden die tekort schieten, in: *Privacy in het informatietijdperk,* S. Nouwt and W. Voermans (eds.), Den Haag: SDU Uitgevers, 17-30.

[34] Custers, B.H.M. (2018) Profiling as inferred data: amplifier effects and positive feedback loops, in: E. Bayamlioglu, I. Baraliuc, L. Janssens and M. Hildebrandt (eds.), *Being Profiled: Cogitas ergo Sum*, Amsterdam, Amsterdam University Press, p. 112–115.

[35] Lila MacLellan, There are only four types of Facebook users, researchers have found, Quartz, July 12, 2017; https://qz.com/1026914/the-four-types-of-facebook-users-relationship-builders-window-shoppers-town-criers-and-selfies/.

[36] Frederick F Schauer, Profiles, Probabilities and Stereotypes, The Belknap Press of Harvard University Press (2006).

[37] *Barocas, S., and Selbst, A. (2016) Big Data's Disparate Impact, California* Law Review 104(3): 671–732.

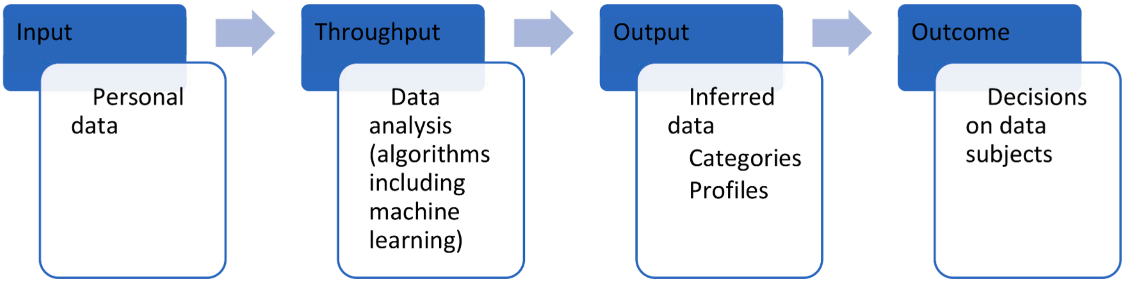
Input
Personal data
Throughput
Data analysis (algorithms including machine learning)
Output
Inferred data
Categories
Profiles
Outcome
Decisions on data subjects

Fig. 1. Different stages in data processing.

data, which means the insurance premium profiles are influenced twice by the original income data: directly and indirectly *via* the inferred credit scores. The reuse of inferred data may thus lead to self-fulfilling prophecies – a phenomenon well-known in profiling.[38] In case of profiling based on inferred data, however, the effect might be much stronger: because of the self-reinforcing effect, patterns may be amplified and become much more entrenched. These effects may amplify inequality, undermine democracy, and further push people into categories that are hard to break out.[39]

For data subjects, it can be important to know which categories a data controller uses and in which category a data subject is placed. The category to which a data subject is assigned can have considerable consequences, ranging from receiving less or no special offers to higher prices or even full exclusion of particular products and services.[40] Hence the categories to which data subjects are assigned strongly influence their reputation from a data controller's perspective.[41] For categories and profiles, awareness and transparency issues are even more complicated than for inferred data. In case of inferred data, data subjects may not be aware of the inferred data. In case of categories and profiles, data subjects may not be aware of the existence of these practices, of the numbers and types of categories and profiles, and of the categories and profiles they are assigned to.

The GDPR acknowledges and addresses these practices of profiling and predictions to some extent, by granting data subjects the right to receive meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject (Articles 13.2.f, 14.2.g, and 15.1.h GDPR). Some have dubbed this as a 'right to explanation',[42] but that is a controversial interpretation. The GDPR also grants the right to object to the use of personal data for *profiling* but only in limited circumstances (Article 21). Importantly, the GDPR does not apply to profiles and categories when they do not fit the definition of personal data, i.e., when these are not personally identifiable information.[43] In such cases, the rights mentioned in the previous paragraph likely do not apply although individuals may still have legitimate reasons for exercising them, for instance, to protect privacy of certain groups of people or to prevent the spread of biases and inaccuracies that they contain.[44]

Although the inferred data and categories and profiles may not always be personal data as defined in Article 4.1 GDPR, in this paper, the focus is inferred data and categories and profiles that are related to identified or identifiable individuals and, therefore, within the scope of the GDPR. However, as we will discuss in Section 3, the fact that these data types can be beyond the scope of the GDPR is an important limitation of the GDPR's scope, and, on the flip side, a reason why controllers prefer to work with de-identified data (mostly to avoid privacy compliance issues).

3. Data subject rights

The EU data protection reform in 2016 introduced an extended section on data subject rights, strengthening data subject rights and empowering data subjects, i.e., giving them more control over their personal data, in order to offer some leverage against powerful data controllers. Under the GDPR regime, the catalogue of control rights consists of the following entitlements: the right to information (Articles 12 through 14), the right to access (Article 15), the right to rectification (Article 16), the right to erasure/to be forgotten (Article 17), the right to restriction of processing (Article 18), the right to data portability (Article 20), the right to object, and the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning an individual or similarly significantly affects her (Articles 21 and 22). In this section, each of these rights is examined in the light of the different types of personal data identified in the previous section.

3.1. *The right to information (Art. 12-14 GDPR)*

The right to information can be considered the cornerstone of the system of data subject rights in the GDPR. It is *primus inter pares* among the data subject rights. Formally, all the rights are deemed equal, but in

[38] A typical example of such a self-fulfilling prophecy is hot spots profiling in policing. If police surveillance mostly take place in neighbourhoods with ethnic minorities, it is not surprising that police databases get filled with data on ethnic minorities. When profiling is based on such datasets, the result may be that these neighborhoods are identified as hot spots for crime. Subsequently, the police is likely to use these risk profiles as a basis for focusing their surveillance on these hot spots. Cf. Custers, B.H.M. (2013) Data Dilemmas in the Information Society. In: Custers B.H.M., Calders T., Schermer B., Zarsky T. (eds.) *Discrimination and Privacy in the Information Society*. nr. 3 Heidelberg: Springer.

[39] O'Neil, C. (2016) *Weapons of Math Destruction; How big data increases inequality and threatens democracy*. New York: Crown.

[40] Poort, Joost and Zuiderveen Borgesius, Frederik, Personalised Pricing: The Demise of the Fixed Price? (February 25, 2021). Available at SSRN: https://ssrn.com/abstract=3792842 or https://doi.org/10.2139/ssrn.3792842.

[41] Solove, D.J., 2007, The Future of Reputation, New Haven, Yale University Press.

[42] Wachter, S., Mittelstadt, B., Floridi, L., 'Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation', *International Data Privacy Law*, Vol. 7, No. 2, 2017, pp. 76–99.

[43] Cf. Finck, M., & Pallas, F. (2020). They who must not be identified—distinguishing personal from non-personal data under the GDPR. *International Data Privacy Law, 10*(1), 11-36; van der Sloot, B. (2020). Regulating non-personal data in the age of Big Data. *Health Data Privacy under the GDPR. Routledge. Num Pages, 21*; Irti, C. (2022). Personal Data, Non-personal Data, Anonymised Data, Pseudonymised Data, De-identified Data. *Privacy and Data Protection in Software Services*, 49–57.

[44] For the importance of protecting group privacy from harms that are not tied to abuses of personal information, see: Galič, Maša, Smart Cities as 'Big Brother only to the Masses': The Limits of Personal Privacy and Personal Surveillance (September 5, 2022). Surveillance & Society 2022, Available at SSRN: https://ssrn.com/abstract=4228444 or https://doi.org/10.2139/ssrn.4228444; Group Privacy: the Challenges of New Data Technologies, Eds. Taylor, L., van der Sloot, B., Floridi, L. Springer: 2017.

practice the right to information stands out as it exemplifies the principle of transparency and represents the focal point for all other data subject rights.[45] Without the necessary information, a data subject cannot meaningfully participate in the data economy, nor can she exercise her other control rights.[46]

The right to information addresses several types of information beyond the actual personal data that a data controller is processing. For instance, it includes information on the identity of the data controller, contact details of the data protection officer, the purposes of and legal basis for the processing, the recipients of the personal data, and storage times.[47]

Since the right to information does not address the actual personal data that is being collected and processed, it is difficult to match it to the types of personal data mapped in the previous section. However, the right to information does provide some points of reference to this taxonomy. First, the distinction between personal data collected directly versus indirectly from the data subject is clearly marked by the two separate articles the GDPR contains for these situations. Hence, Article 13 GDPR applies to personal data collected directly from the data subject and Article 14 GDPR applies to personal data collected indirectly from the data subject (i.e., from a third party). This leaves open the question how the GDPR addresses inferred data and profiles in the context of the right to information. The only indication in Articles 12-14 of the GDPR relating to inferred data and profiles are the right to information on the purposes for which the processing takes place (which may provide clues about what kind of data is inferred and what kind of profiles may be used)[48] and the right to receive meaningful information about the logic involved if automated decision-making, including profiling, is used (which may provide clues about how profiles are established).[49]

The right to information in Articles 13-14 GDPR applies to personal data collected directly from the data subject and personal data collected indirectly from the data subject respectively. However, article 13 GDPR cannot apply to inferred data and categories and profiles in which a data subject is placed, since the data subject does not have such information and therefore cannot provide it. Both Article 13 and 14 contain provisions that data controllers need to provide information about the purposes for which personal data is processed and information about the logic involved in profiling practices, but this does not mean they need to provide information on inferred data or profiles (at least not on the basis of these articles, using grammatical interpretation). Data subjects can, however, derive from Articles 13.2.f, 14.2.g GDPR a right to receive information on profiling. This right to meaningful information about the logic of the profiling may not entail information about the actual profiles or the categories in which data subjects are placed, but it may provide clues about what kinds of profiles or categories are established. Some have argued that in the case of inferred data, profiles or categories, Article 13 (and likely Article 14 as well) of the GDPR is essentially not triggered, because this is neither data obtained directly nor indirectly from the data subject – it is data generated by the data controller.[50] However, inferred data and profiles could be interpreted as 'self-obtained' by data controllers or could be obtained from other data controllers, in which case it could be possible to argue that they fall under the definition of Article 14. A broader, and indeed teleological, interpretation of the provision is also supported by the Court of Justice of the EU (CJEU) decision in SCHUFA where it stated, in passing, that the rights, including the right to information, should be interpreted "by the purpose pursued by Article 22 of the GDPR, consisting of protecting individuals against the particular risks to their rights and freedoms represented by the automated processing of personal data, including profiling."

3.2. *The right of access (Art. 15 GDPR)*

Contrary to the right to information, which aims to facilitate control mostly in the stage *before* data processing starts, the right of access applies in *subsequent* stages of data processing. The right of access is to some extent also complementary to the right to information: whereas the right to information addresses procedural information on who is processing the data and for which purposes, the right of access addresses substantive information on which personal data is actually being processed. In this respect, Article 15.3 GDPR is relevant, which states the right of data subjects to receive a copy of the personal data that is being processed.

The right of access, as one of the control entitlements, represents a key element in enhancing user control over their personal data.[51] The right entitles a data subject to receive information on whether or not her personal data is being processed, and if so, to access her personal data including additional information about data processing.[52] The objective of the right is to provide comprehensive access to data about an individual's use of a service, conveniently, securely, privately, and free of charge.[53]

It is obvious that Article 15 applies to both information collected directly and indirectly from data subjects. If the right of access does not cover this type of information, then what does it cover? When data subjects request a copy of their personal data that is being processed, it is most likely that this contains information directly obtained from data subjects. Given the phrasing in Article 15.1 GDPR, the right of access also contains the right to obtain confirmation as to whether or not a data controller is processing the data subject's personal data. For example, some people are not Facebook members, but nevertheless make use of Facebook's public pages or 'like' plug-ins when they surf other websites. Facebook also processes these persons' personal data (such as IP addresses).[54] Such personal data, obtained indirectly from the data subject is in scope of the right of access because Article 15 GDPR allows for the

[45] It has also been referred to as 'vital', at least in certain contexts such as the gig economy. Naudts, Laurens and Dewitte, Pierre and Ausloos, Jef, Meaningful Transparency through Data Rights: A Multidimensional Analysis (July 27, 2021). (Forthcoming), in Eleni Kosta, Ronald Leenes and Irene Kamara (eds), Research Handbook on EU data protection, Edward Elgar, 2022, Amsterdam Law School Research Paper No. 2022-37, Institute for Information Law Research Paper No. 2022-02, page 3, https://ssrn.com/abstract=3949750.

[46] Helena U. Vrabec, Data subject rights under the GDPR, Oxford University Press, 2021, p. 64.

[47] Article 13 (1) and (2)(a) GDPR.

[48] See Article 29 Working Party: Guidelines on transparency under Regulation 2016/679, p. 14, fn. 30 (as last revised and adopted on 11 April 2018): "If the purpose includes the creation of inferred personal data, the intended purpose of creating and further processing such inferred personal data, as well as the categories of the inferred data processed, must always be communicated to the data subject at the time of collection, or prior to the further processing for a new purpose".

[49] Article 13(2)(g) and 12(2)(f) GDPR.

[50] Wachter, Sandra and Mittelstadt, Brent, A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI (October 5, 2018). Columbia Business Law Review, 2019(2), p. 545.

[51] European Commission, 'Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: A Comprehensive Approach on Personal Data Protection in the European Union' (2010) 7.

[52] The latest CJEU case law interpreted the right broadly, confirming that it includes the provision of the identity of specific recipients of personal data if the data subject requests it. C-154/21 RW v Österreichische Post.

[53] Simone Fischer-Hübner and others, 'Online Privacy: Towards Informational Self-determination on the Internet' in Mireille Hildebrandt and others (eds), Digital Enlightenment Yearbook 2013 (2013) 133.

[54] The so called shadow profiles; see for instance Gennie Gebhart, 'Facebook, This is not what "complete user control" looks like' (*Electronic Fontier Foundation*, 11 April 2018) <https://www.eff.org/deeplinks/2018/04/facebook-not-what-complete-user-control-looks> accessed 7 June 2018.

option to obtain confirmation from the data controller.

It is clear that the scope of the right of access reaches beyond the information data subjects themselves previously provided. Recital 63 of the GDPR provides several examples of the information within the scope of the right of access, including data in medical records containing information such as diagnoses, examination results, assessments by treating physicians and any treatment or interventions provided. Typically, these types of information are generated by doctors rather than patients, meaning that inferred data is indeed within the scope of Article 15 GDPR.[55]

However, it is not clear how far this extends. Some inferred data may relate to groups and, as long as it is not ascribed to individual group members, may not meet the definition of personal data. For instance, if the data in a database reveal a pattern that only adults have a driving license, this is not necessarily personal data. As long as it is unknown whether John has a driving license, it cannot be concluded that John is an adult. Only when it becomes known that John has a driving license, the characteristic that he is an adult can be ascribed to him, which then qualifies as personal data.

Even if the inferred data is personal data in the sense that the inferences are ascribed to individuals, it may be unclear whether the right of access can be used by data subjects to get access to such information. This is due to two types of reasons.

First, the CJEU has taken a strict view as to the scope of personal data when the data is being used in an analytical process. In *YS and others*[56], the Court followed the opinion of Advocate General Sharpston in which she stated that: "*Only information relating to facts about an individual can be personal data. […] a legal analysis is not such a fact. Thus, for example, a person's address is personal data but an analysis of his domicile for legal purposes is not*".[57] In *Nowak*[58], the court expanded its view as it held that examiners' notes (but not exam questions!) constitute personal data.[59] Applying these two decisions to inferred data, one could argue that the input data and inferred data are personal data and hence within the GDPR scope, but the model based on which the inference is made, is not and cannot be accessed.[60] However, it should be noted that these decisions were focused on non-digital environments and the court has not yet had a chance to assess the application of the right to inferred data in a digital, algorithm-run environment. Given that the judgements appear slightly outdated and not necessarily on point, it is hard to concede the argument that they should be considered authoritative for the applications of data subject rights to inferred data. Whether this is a correct position to hold or not might soon be clarified as a preliminary question of an Austrian court was recently filed with the CJEU and indicates this precise matter.[61]

Second, there are various exceptions that apply to the right of access. For instance, inferred data may constitute or reveal trade secrets: if a data controller regularly reveals inferred data, it may become possible, *via* reverse engineering, to disclose how their analyses and software work.[62] Article 15.4 GDPR provides an exemption to the right of access, stating that a copy of the data does not have to be provided if that adversely affects the rights and freedoms of others (in this case the rights of the data controller). Recital 63 further explains that the right of access should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software.[63] Providing complete information about the processing of inferred data may also be challenging from a practical point of view, as the data may be difficult to capture across the internal systems and display it in a meaningful way. When data subject requests trigger searches that are too onerous for data controllers, the courts might find those requests disproportionate. Currently, balancing the right of access with competing interests such as trade secrets is to a large extent the discretion of the data controller. There is hardly any case law in this area.[64] Hence, much depends on the willingness of data controllers when it comes to providing access to inferred data upon access requests. That said, the EDPB signaled their preference for fully accommodating access to profile data by recommending to data controllers to enable an automated process for data subjects to access their profile information.[65]

Article 15.1.h GDPR contains a similar provision as Articles 13.2.f and 14.2.g, stating that data subjects have a right to meaningful information about the logic involved in case profiling is used (and, preferably, also when other types of processing are used).

This provision may not entail information about the actual profiles or

[55] EDPB, Guidelines 01/2022 on data subject rights - Right of access, 18 January 2022, page 33. Also see Custers, B.H.M., 'Profiling as inferred data: amplifier effects and positive feedback loops', in: E. Bayamlioglu, I. Baraliuc, L. Janssens and M. Hildebrandt (eds.), *Being Profiled: Cogitas ergo Sum*, Amsterdam, Amsterdam University Press, 2018, pp. 112-115; Custers, B.H.M. (2005) The risks of epidemiological data mining. In H. Tavani (Ed.), *Ethics, computing and genomics: Moral controversies in computational genomics*. Boston: Jones and Bartlett Publishers Inc.

[56] Joined Cases C-141/12 and C-372/12, *YS (C-141/12) v Minister voor Immigratie, Integratie en Asiel, and Minister voor Immigratie, Integratie en Asiel (C-372/12) v M and S,* ECLI:EU:C:2014:2081, para 48.

[57] Opinion of Advocate Generale Sharpston delivered on 12 December 2013 in joined cases Joined Cases C-141/12 and C-372/12, *YS (C-141/12) v Minister voor Immigratie, Integratie en Asiel, and Minister voor Immigratie, Integratie en Asiel (C-372/12) v M and S,* ECLI:EU:C:2014:2081.

[58] Case C-434/16, *Peter Nowak v Data Protection Commissioner***,** ECLI:EU:C:2017:994.

[59] Davinia Brennan, The expanding scope of 'personal data' – CJEU delivers judgment in Nowak, Ireland IP & Technology Law Blog, 2 January 2018, accessible at: https://www.irelandip.com/2018/01/articles/cyber-risk-data-privacy/expanding-scope-personal-data-cjeu-delivers-judgment-nowak/.

[60] *Supra* 27.

[61] Court of Justice of the European Union. "Request for a preliminary ruling from the Verwaltungsgericht Wien (Austria) lodged on 16 March 2022 – CK (Case C-203/22)", available at: https://curia.europa.eu/juris/document/document.jsf?text=algorithm&docid=260303&pageIndex=0&doclang=en&mode=req&dir=&occ=first&part=1&cid=8960564#ctx1.

[62] Hildebrandt, M. (2011) *The Rule of Law in Cyberspace?* Inaugural Lecture, Nijmegen: Radboud University. https://works.bepress.com/mireille_hildebrandt/48/.

[63] Brkan also points out that state secrets could be an obstacle to algorithmic transparency. Brkan, Maja, Do Algorithms Rule the World? Algorithmic Decision-Making in the Framework of the GDPR and Beyond (August 1, 2017). A revised version of this paper has been published in International Journal of Law and Information Technology, 11 January 2019, DOI; 10.1093/ijlit/eay017, Available at SSRN: https://ssrn.com/abstract=3124901 or https://doi.org/10.2139/ssrn.3124901. Interestingly, the EDPB guidance on the general restrictions to data subject rights in Article 23 of the GDPR neglects the rights of data controllers such as trade secrets and interprets the exception narrowly - Guidelines 10/2020 on restrictions under Article 23 GDPR, October 13, 2021. It does, however, recognize those controllers' rights as attached to Article 15 of the GDPR. EDPB, Guidelines 01/2022 on data subject rights - Right of access, 18 January 2022, page 52.

[64] Some exceptions are *Bundesgerichtshof*, 28 January 2014, VI ZR 156/13, https://openjur.de/u/677956.html; Dutch Supreme Court, 17 August 2018, ECLI:NL:HR:2018:1316. This topic was also left out of the court decision in *SCHUFA*, although the AG Opinion does deal with the interpretation of trade secrets as an exception to the data subject access right: "*It follows that while protection of trade secrets or intellectual property in principle constitutes a legitimate reason for a credit information agency to refuse to disclose the algorithm used to calculate the score for the data subject, it cannot under any circumstances justify an absolute refusal to provide information, a fortiori where there are appropriate means of communication which aid understanding while guaranteeing a degree of confidentiality.***"(**Recital 56).

[65] "Controllers may want to consider implementing a mechanism for data subjects to check their profile, including details of the information and sources used to develop it." Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' [2017] 19.

the categories in which data subjects are placed, but it may provide clues about what kinds of profiles or categories are established. This provision does not, however, entail a right for data subjects to receive information about the actual profiles or categories, such as the different types of categories and the category in which the data subject is placed. The provision in Article 15.4 GDPR (stating that all this shall not adversely affect the rights and freedoms of others) could considerably restrict data subject access to categories and profiles. On the one hand, providing such information may interfere with the interests of the data controller, as such information may be considered trade secrets that provide a data controller with competitive edge. It may also create excessive business cost for the data controller to start providing details of how the data has been managed on their end, through which models it has gone and what the analysis has looked like.[66] On the other hand, providing such information may interfere with the interests of other data subjects in the database, as such information may also reveal their characteristics. Particularly if a data subject raises a question like “who else is in my category?”, this cannot be answered at an individual level.

3.3. The right to rectification (Art. 16 GDPR)

According to Article 16 GDPR, the data subject has the right to have inaccurate personal data concerning him or her rectified. Also, the data subject has the right to have incomplete data completed, for instance, by means of providing a supplementary statement.

For data that was obtained directly or indirectly from the data subject, this provision is relatively straightforward. This often concerns personal data of a rather factual nature. An address is either correct or incorrect. An address can be incorrect, for instance, because it contains a typo or is outdated. Labeling someone as a terrorist may be inaccurate because of a mistake in the process of mapping the individual’s name with the OFAC’s list. Such inaccurate predictions may lead to traumatic experiences. For instance, in a US case *TransUnion v. Ramirez*, the lead plaintiff only learned about the error of being mislabeled a terrorist when he tried to buy a car. He was stunned after the car salesperson informed him the purchase was impossible because he was on a terrorist list.[67]

Also for more subjective personal data, such as a person’s indicated preferences or interests, rectification is straightforward. Someone’s preferences or interests may change over time and data subjects can inform data controllers about this. The data controller may not be in a position to check whether the claim of the data subject is correct and may have to assume this is the case. Rectifying and updating personal data generally is in the interest of both the data controller and the data subject. For the data controller, implementing rectifications and updates is also an obligation according to Article 5.1.d GDPR, which states that personal data shall be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay.[68]

For inferred data and categories and profiles, this is much more complicated. Since data analytics often makes use of statistics, these types of personal data often contain error margins. For instance, from someone’s online behavior her music preferences may be inferred. Even if this can be done rather accurate, say with 95 % accuracy, the accuracy is not perfect. According to the text of Article 16 GDPR, a data subject can argue that this piece of inferred data is (factually) incorrect and needs to be rectified. At the same time, however, the data controller can argue that this piece of inferred data is (statistically) correct, inferred *via* sound statistical methods and perhaps even labeled with metadata that indicates the error margins. It is unclear which of these interpretations prevails.[69] It could be argued that both interpretations have their merits and therefore a data subject should be allowed to provide a supplementary statement, so that both views coexist in the database. However, according to the text of Article 16 GDPR, the possibility of providing a supplementary statement is only allowed for completing incomplete personal data, not for rectifying inaccurate personal data (or, even less likely, rectifying the models that learn from non-personal data).

In terms of inferred preferences, a data subject can indicate whether the resulting characteristics ascribed to her are correct or not. For some inferred characteristics this is basically impossible. Typically, for predicted characteristics a data subject is not in the position to assess whether these are correct. If a data controller assesses that a data subject has 75 % probability of attracting cancer in the next five years, the data subject cannot assess whether that is correct, not even after the five years have expired. The same applies to life expectancies or credit scores. Instead of assessing whether the outcomes are correct, a data subject could perhaps assess whether the characteristics were inferred correctly. However, that would involve access not only to the data subject’s own personal data, but also access to the personal data of others involved in the analyses and access to the data analytics tools used. This would immediately interfere with the rights and freedoms of others and therefore not allowed under the GDPR.[70] In some cases, a data subject and a data controller may have different opinions on whether data are correct. For instance, on the basis of online behavior and using sophisticated data analytics a data controller may infer the preferences of the data subject. These may not be the same as preferences the data subject recently indicated herself. The data subject may not be aware (yet) of some of her preferences or may not recognize these preferences, whereas the data controller does. Basically, the data controllers then know the data subject better than she knows herself. In such cases, the data subject may try to invoke the right to rectification, but it is unclear whether the data controller can overrule such requests on the basis of the argument that the data is correct even though the data subject thinks otherwise. This is why some have argued that whether data is verifiable should not be the limitation to the right of rectification – as long as processing of data has an impact on the individual, it does not really matter, from the privacy harm perspective, whether it can be verified or not.[71] From a practical perspective, this point of view remains problematic, however, since this would not offer clarity on when data that cannot be verified should be rectified. For unverifiable data, the right to erasure perhaps has more to offer for data subjects than the right to rectification.[72]

Beyond the question of verifiability, there is a further issue with the right of rectification. To effectively rectify inferred personal data, one needs to be able to assess and rectify the accuracy of the logic which was used to draft those inferences. The CJEU case law does not seem to support the extension of the right of rectification to the data analytics

[66] Case C-434/16, *Peter Nowak v Data Protection Commissioner*, ECLI:EU:C:2017:994.

[67] TransUnion LLC v. Ramirez, 141 S.Ct. 2190 (2021). Solove, 2022.

[68] For a more detailed analysis of the controller’s obligations with regards to data accuracy see: Dimitrova, Diana, The Rise of the Personal Data Quality Principle. Is it Legal and Does it Have an Impact on the Right to Rectification? (February 10, 2021), https://ssrn.com/abstract=3790602 or https://doi.org/10.2139/ssrn.3790602.

[69] See Staunton, C. (2021). Individual Rights in Biobank Research Under the GDPR. In: Slokenberga, S., Tzortzatou, O., Reichel, J. (eds) GDPR and Biobanking. Law, Governance and Technology Series, vol 43. Springer, Cham. https://doi.org/10.1007/978-3-030-49388-2_6 for a view that uncertain data cannot be captured by the right to rectification (in the context of genome data used by biobanks). “Uncertainty does not equate to inaccuracies, and biobanks will only be required to update any inaccurate information”.

[70] Jove, Daniel: Peter Nowak v Data Protection Commissioner: Potential Aftermaths Regarding Subjective Annotations in Clinical Records, European Data Protection Law Review 2019 Vol.5 N°2 p.183.

[71] Wachter, *supra* 50, referencing Kaumann and Braun, pp. 57–58.

[72] Which is actually what the Court in Nowak seems to suggest. Nowak, para 55.

tools and personal data of others involved in the analyses.[73] For example, exam notes containing assessments or opinions that do not constitute personal data cannot be challenged under the existing mechanisms of data protection law.[74] Applying this to the context of inferred data, such logics behind the inferred data cannot be rectified under data protection law and can only be contested if there is a procedure in place to contest the evaluation.[75] In essence, a data subject cannot assess whether inferred data, predictions, categorizations, and profiles are correct and therefore is blocked from effectively invoking the right to rectification. This is a major limitation of the right to rectification.[76]

Another limitation is that the data subject's right to provide a supplementary statement to the data controller in order to complete incomplete data will not make much of a difference. Such a statement can be included in the databases, but it is unlikely to be included in any data analytics. The current state of the art of technology usually does not allow for including such statements, which usually cannot be included in the relational databases, but are stored elsewhere, in a different part of the IT system or even a separate IT system.[77] Article 16 GDPR does not require data controllers to include the supplementary statements in subsequent analyses or otherwise take them into account.

3.4. *The right to erasure (Art. 17 GDPR)*

Data subjects can invoke Article 17 of the GDPR to have personal data removed, including inferences,[78] in case their consent is successfully withdrawn, or in case they have a stronger interest than the legitimate interest of the data controller, allowing the processing to be stopped. However, the right to erasure, like some other provisions of data protection law we discussed above, appears to lose its strength when it applies to inferred data. This is because of several technical and economic factors present in the current data markets, such as the use of AI and ubiquitous data sharing, that influence the use of inferred data.[79]

The first problem is that the nature of inferred data dilutes a data subject's legitimate interest in the inferred data in comparison to that of a data controller.[80] Contrary to 'ordinary personal data' such as names and email, inferred data are likely to be seen as 'owned' by a data controller since they are intimately connected with the model that the data controller constructed.

The second problem is that the inferred data might have been shared with third parties (new controllers) in which the GDPR does not require deletion, but only informing other controllers of the data subject request.[81] This light-weight requirement comes with some exceptions, notably that the controller may avoid it if the burden imposed is disproportionate. But even when a request *is* proportionate, the system lacks accountability since no deletion is required when a data controller is not in a position to identify the data subject.[82] This is commonly the case when the data controller has de-identified the personal data for analytical purposes. The controller does not need to re-identify the data in order to allow the data subject to exercise his or her rights.[83]

The third problem is that the scope of the data protection law is attached to the concept of identifiability,[84] which creates a gap in the application of the right to erasure. Without having a profile or a model applied to a data subject, data protection law (and in turn data subject rights) do not apply. Of course, having the opportunity to achieve deletion after an inference has been used to identify a data subject and/or applied to her is most critical, but there is value in having some control over that de-identified model or a profile as well. Depersonalized models might be discriminatory or incorrect, and without a data protection recourse, a data subject might continue to see similarly problematic decisions being taken in the future.[85] However, if a data subject alone requests erasure, it is unlikely that withdrawing one person's data will make much difference to a trained model and/or the algorithmic outcome. To make effective use of the right to erasure to alter models, whole groups would need to collaborate explicitly or implicitly to request erasure, which is highly unlikely.[86]

The fourth problem is related to this: it is increasingly easy to predict any missing data in datasets.[87] This means that if data is erased on the request of a data subject, it may still be possible to infer it again. Obviously, it can be argued that the data controller is not allowed to do this, but automated data analytics tools may keep doing this despite legal restrictions. A typical example of this are the periodic updates that some databases get from other databases. In case a master database from the central government updates the decentralized databases of local municipalities every night. If a data subject requests erasure of some data at the municipality and the municipality manually removes the data, it will be back in their database the next day. It is sometimes argued that the law does not keep up with technology, but in such situations, it could also be argued that the technology does not keep up with the law. The technology could also be designed in other ways, avoiding interference with data subject rights.

3.5. *The right to restriction of processing (Art. 18 GDPR)*

Article 18 GDPR allows data subjects to obtain from the data controller restriction of the processing. Restriction of processing is defined in Article 4.3 GDPR as the marking of stored personal data with the aim of limiting their processing in the future. Recital 67 of the GDPR explains that this could include, inter alia, temporarily moving the

[73] Nowak; Case C-141/12 and C-372/12, YS and Others, EU:C:2014:2081.

[74] *Nowak*, para. 54.

[75] Wachter, *supra* 50, p. 59.

[76] See Dara Hallinan, Frederik Zuiderveen Borgesius, Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle, International Data Privacy Law, Volume 10, Issue 1, February 2020, page 3, https://doi.org/10.1093/idpl/ipz025.

[77] Similar technical challenges are present in some other complex technological environments such as blockchain. Voss, W. Gregory, Data Protection Issues for Smart Contracts (June 3, 2021). Smart Contracts: Technological, Business and Legal Perspectives (Marcelo Corrales, Mark Fenwick & Stefan Wrbka, eds., Hart Publishing/Bloomsbury, 2021), https://ssrn.com/abstract=3977477.

[78] Wachter, *supra* 50, 62.

[79] For a helpful overview of the challenges to the right to erasure (and some other data subject rights) in the AI context see: Zhang, D., Finckenberg-Broman, P., Hoang, T., Pan, S., Xing, Z., Staples, M. and Xu, X., 2023. Right to be forgotten in the era of large language models: Implications, challenges, and solutions. arXiv preprint arXiv:2307.03941.

[80] Wachter, *supra* 50, fn. 393.

[81] Article 17 in connection with Article 19 of the GDPR.

[82] Article 11(2) GDPR. Stalla-Bourdillon, Sophie and Knight, Alison, Anonymous Data v. Personal Data — A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data (March 6, 2017). Wisconsin International Law Journal, 2017, p. 19, https://ssrn.com/abstract=2927945.

[83] Wachter, *supra* 50, 65.

[84] Erdos, David, Identification in EU Data Protection Law (January 6, 2021). University of Cambridge Faculty of Law Research Paper No. 4/2021, https://ssrn.com/abstract=3761068. A similar stance is taken by Mantelero: Alessandro Mantelero, Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection, Computer law & security review 32 (2016) 246.

[85] In a similar vein, it has been argued that anonymised datasets can still present residual risks to data subjects. Jef Ausloos, Michael Veale and René Mahieu, Getting Data Subject Rights Right, 10 (2019) JIPITEC 283, 295. As best practice with regards to applying the right to delete in the context of machine learning, Hawkings et al. suggest that 'exact erasure' is used. Exact erasure means the deletion of the data in the model *and* in the training dataset.

[86] Vrabec (2021) 223.

[87] Hond, A. A. H. de. (2023, October 11). From code to clinic: theory and practice for artificial intelligence prediction algorithms, https://hdl.handle.net/1887/3643729.

selected data to another processing system, making the selected personal data unavailable to users, or temporarily removing published data from a website.

The right to restriction of processing can only be invoked in some situations: if the accuracy of the personal data is contested by the data subject, if the processing is unlawful, if the data controller no longer needs the data but the data subject does, or if the data subject has invoked Article 21 GDPR (which is discussed below in Section 3.7).

In case of contested accuracy, the situation is similar to that of Article 16 GDPR. If the personal data are obtained directly or indirectly from the data subject, this can be applied straightforwardly. But if the personal data concern inferred data, predictions, categorizations or profiles, the data subject is not in the position to assess accuracy and Article 18 GDPR will be hard to invoke on this basis. Similar to invoking the right to rectification, verifiability may be a complicating issue.

In case of unlawful processing, the data subject has the option to invoke the right to erasure or the right to restrict the processing. Opting for the latter seems to make sense if the data subject expects she may want to resume the processing at a later stage or may still need the personal data for other purposes, such as the establishment, exercise, or defense of legal claims, which resembles the next situation listed in Article 18 GDPR.[88]

In case the data controller no longer needs the personal data, but the data subject does, the processing can also be restricted. In this situation, like in the case of unlawful processing, any blocking of the processing applies in principle to all types of personal data that are being processed, including inferred data and profiles.[89]

In summary, Article 18 GDPR can be applied straightforwardly to personal data obtained directly or indirectly from the data subject. Applying it to inferred data and categorizations and profiles is more complicated: in case of contested accuracy this is not possible, but in case of unlawful processing or obsolete data it can be applied. However, in these cases the right to erasure may be preferred by the data subject, making the right to restrict the processing somewhat redundant.[90] The only practically feasible use of Article 18's right in the context of inferred data is therefore in connection with the right to object (Article 21), particularly in relation to the processing of data for marketing purposes.[91]

3.6. The right to data portability (Art. 20 GDPR)

The right to data portability in Article 20 of the GDPR allows data subjects to receive personal data that they have previously provided to data controller in a format that enables frictionless porting of the data to another controller. Certain categories of data such as name, email, and personal address fall squarely within the scope of the portability right as they are clearly provided by data subjects. Similarly, data provided *about* the person, for instance, location data or log-ins from visited websites, are covered by the right as well.[92]

However, once data has been analyzed using any sort of algorithmic techniques to draw useful insights (ie, inferred data), the results of this analysis should not be ported.[93] It is arguable that in applying analytical techniques, data loses the direct connection with data subjects and is thus no longer considered to be 'provided by them'.[94] Rather, it concerns data generated by the data controller. A user's profile created using the analysis of raw smart metering is one such example. Some types of data may fall between raw data and derived data, such as reputation scores that are attained by users of online marketplaces such as Airbnb.[95] If the scores were portable, this would mean that Airbnb users would have the right to take their reviews and transfer them to a competitor, for example, Couchsurfing.[96]

The interpretation of 'provided data' is one of the most contestable aspects of the GDPR's provisions on data portability, yet a critical one as it can open up or close down the portability of a large amount of personal data. Authorities have not yet decided what the boundaries of data portability should be.[97] In fact, the EC criticized the Article 29 Working Party for adopting a position that was too data-subject centric.[98]

Sceptics might argue that the narrow scope is not that much of an issue, as there is a more general, fundamental problem with the data portability right which is that it rests on the assumption that there is a healthy market for privacy which is an unrealistic proposition according to some scholars.[99] This view overlooks the fact that data portability is more than a vehicle of competition. The right to data portability also functions as a vehicle of transparency and fairness.[100] In connection with these two functions, the exclusion of inferred data from its scope is concerning, as it means less information about the building blocks of the data economy.

3.7. The right to object and automated decision-making (Art. 21-22 GDPR)

Article 21 of the GDPR allows a data subject to file an objection with the data controller and request to stop (restrict) the processing of her personal data if required by the particularities of a situation that she finds herself in. This right is absolute when personal data is used for marketing purposes, especially for profiling.[101] The absolute nature of the right in cases of profiling is understandable and should not come as a shock for the industry. Profiling in marketing campaigns may come

[88] Taner Kuru, Iñigo de Miguel Beriain, Your genetic data is my genetic data: Unveiling another enforcement issue of the GDPR, Computer Law & Security Review, Volume 47, p. 11, 2022.

[89] Kaltheuner, F. and Bietti, E., 2017. Data is power: Towards additional guidance on profiling and automated decision-making in the GDPR. *Journal of Information Rights Policy and Practice*, 2(2), p.16.

[90] Which, perhaps, contributed to the right being underused by data subjects. Ausloos, Veale and Mahieu report that the right has been ignored by most data controllers. Jef Ausloos, Michael Veale and René Mahieu, Getting Data Subject Rights Right, 10 (2019) JIPITEC, p. 283.

[91] *Ibid*, p. 301.

[92] EDPB, Guidelines on the right to data portability, WP 242 rev.01, p. 10, https://ec.europa.eu/newsroom/article29/items/611233.

[93] *Ibid.*

[94] *Ibid.*

[95] Zanfir-Fortuna, Gabriela, The Right to Data Portability in the Context of the EU Data Protection Reform (March 12, 2012). International Data Privacy Law, Vol. 2, No. 3, 2012, p.3, https://ssrn.com/abstract=2215684.

[96] Helena Ursic, Unfolding the New-Born Right to Data Portability: Four Gateways to Data Subject Control, SCRIPTed, Volume 15, Issue 1, August 2018.

[97] For a good overview of the discussions around the scope of the right see Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, Ignacio Sanchez, The right to data portability in the GDPR: Towards user-centric interoperability of digital services, Computer Law & Security Review, Volume 34, Issue 2, 2018, Pages 193–203.

[98] David Meyer, 'European DPAs Mull Strategy for Tackling Uber's Data Catastrophe' *IAPP Privacy Advisor* (25 April 2017), https://iapp.org/news/a/european-commission-experts-uneasy-over-wp29-data-portability-interpretation-1/.

[99] Daniel Solove, The Limitations of Privacy Rights, GW Law Faculty Publications & Other Works (2022), page 29. Also see Jurre Reus and Nicole Bilderbeek, Data portability in the EU: An obscure data subject right, Mar 24, 2022, The IAPP Privacy Perspectives, https://iapp.org/news/a/data-portability-in-the-eu-an-obscure-data-subject-right/.

[100] Urquhart, Lachlan and Sailaja, Neelima and McAuley, Derek, Realising the Right to Data Portability for the Domestic Internet of Things (March 15, 2017); p.1, https://ssrn.com/abstract=2933448.

[101] Gabriela Zanfir-Fortuna, 'Article 21 Right to Object' in Christopher Kuner and others (eds), The EU General Data Protection Regulation (GDPR): A Commentary (Oxford University Press 2020) accessed 27 March 2023.

across awkward and opaque.[102] From a policy perspective, there is value (and little risk) in granting additional protection against a phenomenon that is omnipresent and arguably lacks legitimacy in the absence of a data subject's consent.[103] Article 21(5) of the GDPR provides strengthened protection by ensuring technical avenues ("by automated means") for data subjects to object. This essentially grants the right to object to the users without requiring them to use legal avenues to request it. For instance, Google allows its users to easily turn off the personalized ads function which relies on the use of profiling techniques.[104]

Although this right clearly applies to inferred data (as profiling is explicitly mentioned), its scope is limited unless direct marketing is involved. A data subject can only object to the processing of her personal data if the processing has been based on a public interest or on a data controller's private legitimate interest. Typically, a data controller can circumvent the right to object by demonstrating a legitimate interest to carry on the processing. Given potential challenges to the right of object in practice, it is worth considering whether the scope of the 'absolute' right to object should be perhaps broadened to include some additional types of data processing, particularly inferred data processing. In an opinion from 2014, the Article 29 Working Party argued for a more widely applicable mechanism for opt-out and easier ways to exercise it in order to satisfy the rights and interests of the data subjects.[105]

Article 22 of the GDPR provides the right not to be subject to automated decision-making, which is similar to the right in Article 21 because, despite of its name, it also functions as some sort of an objection (i.e., restricting) right. Specifically, the Article 22 right focuses on the automatic processing of personal data that produces decisions with legal effects on people and poses some limits to such decisions.

Article 22 functions as a gatekeeper, a blocking mechanism, and can be read in two different ways: either as a right that the data subject can exercise, or as a prohibition for data controllers.[106] The latest opinion by the Article 29 Working Party and the CJEU judgment in *SCHUFA* made it clear that the provision has a prohibitory nature.[107] Data subjects do not need to act to prevent automated decision-making but are rather protected by default.

The GDPR focuses on "automated" decisions, although automation is not really the key feature of what makes certain data-driven decisions problematic.[108] Some have argued that a more apt regulatory focus would be on the use of inference in decisions, as inference involves using existing data to generate new data about a person or to make predictions about them.[109]

To some extent, Article 22 is able to address this. In addition to (generally) prohibiting legally significant automated processing, the article also offers an avenue to control the process of *inferring* personal information, recognizing it as the area where mistakes, biases and inaccuracies can occur that lead to harmful results.

Although Article 22.3 offers data subjects the opportunity to obtain further understanding, this does not necessarily solve the bigger issue which is the prejudicial nature of algorithmic decision-making that is harmful to a larger society not just the individual. Algorithms trained or operated on a real-world data set that necessary reflects existing discrimination may well replicate that discrimination.[110] Algorithms can amplify prejudice in existing data by using it in a widespread systematic way.[111] Providing people with a right to stop solely automated decision-making about them is an insufficient response to these problems, because it does not lead to a better understanding of the nitty-gritty of algorithmic models and/or its safer application.[112]

While the GDPR does not grant the right to investigate any models of data controllers that are used for making inferences,[113] it does provide guidance for controllers by urging them to "*use appropriate mathematical or statistical procedures for the profiling, implement technical and organizational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimized, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject, and prevent, inter alia, discriminatory effects on natural persons.*"[114] However, even this might not be enough according to those who believe that current protections fail to ensure that inferential decisions about people are made fairly, accurately, and consistently with important societal values.[115] "*The law must bring the massive web of inferential decisions about people under control, to address their skewed assumptions, troubling output, amplification of prejudice, and their massive troubling effects on society*."[116] But it seems this is not the mandate that the right under Article 22 has. As a response, some scholars have proposed that this specific area should be tackled by a newly introduced right to reasonable inferences that would draw on general privacy law rather than the identifiability-

[102] Direct marketing relies on creation of general profiles, the so-called personas, which, as previously explained, come with a level of inaccuracy. https://blog.hubspot.com/marketing/buyer-persona-research.

[103] Similarly, from a marketing strategist perspective, it is probably better to skip it all together than risking that customers get upset. For an explanation about why profiling for the purposes of personal advertising cannot in principle rely on legitimate consent see Judgment of the Court in Case C-252/21, Meta Platforms and Others (General terms of use of a social network).

[104] https://support.google.com/My-Ad-Center-Help/answer/12155656?hl=en&co=GENIE.Platform%3DAndroid (accessed February 9, 2024).

[105] 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, 844/14/EN WP 217, 9 April 2014.

[106] Under the former EU Data Protection Directive, which preceded the GDPR and contained a highly similar provision, the authorities' views on its interpretation differed (EU Directive 95/46/EC). Wachter S, Mittelstadt B, and Floridi L, 'Why a Right to Explanation of Automated Decision-making Does Not Exist in the General Data Protection Regulation' [2017] International Data Privacy Law 76, p. 94.

[107] Article 29 Data Protection Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' [2017] 19.

[108] Solove, *supra* 99, p.48.

[109] *Ibid.*

[110] Anupam Chander, The Racist Algorithm?, 115 Mich. L. Rev. 1023 (2017), p. 1036.

[111] Solove, *supra* 99, p. 49.

[112] *Ibid.* Also see Aloisi, Antonio and Gramano, Elena, Artificial Intelligence Is Watching You at Work. Digital Surveillance, Employee Monitoring, and Regulatory Issues in the EU Context (June 10, 2019). Special Issue of Comparative Labor Law & Policy Journal, "Automation, Artificial Intelligence and Labour Protection", edited by Valerio De Stefano, Vol. 41, No. 1, pp. 127, SSRN: https://ssrn.com/abstract=3399548.

[113] See above the discussion on the existing case law.

[114] GDPR, Recital 71.

[115] Solove, *supra* 99, p. 50.

[116] *Ibid.*

dependent data protection law.[117]

4. Analysis

Table 1 provides a summary of the discussion in Section 3, showing to what extent each data subject right encompasses each of the types of personal data distinguished in Section 2. In this section, these findings are further examined and contrasted with a more teleological and contextual approach, considering the intentions and structure of the GDPR.

The findings in Table 1 clearly show that all data subject rights encompass personal data either directly or indirectly obtained from the data subject. In other words, if a data subject provides personal data, all data subject rights fully apply to these data, also when they are transferred to other data controllers. For inferred data, categories, and profiles, this is much less clear. In fact, many data subject rights do not seem to apply to inferred data, categories, and profiles, at least not fully and unambiguously. This is either due to the way they are phrased (limiting their scope), or the exceptions included in the GDPR (limiting their practical value). Only for some rights (i.e., Article 21 and 22) or some provisions added (i.e., Art. 13.2.f, 14.2.g, and 15.1.h) that specifically focus on profiling and automated decisions, the protection offered clearly extends to inferred data, categories, and profiles.

These findings are supported by the textual interpretation of the data subject rights and confirmed in actual practices, in which data controllers often seem not to be inclined to apply data subject rights to inferred data, categories, and profiles. Data controllers have several reasons for this. Practical reasons include time, costs and efforts involved, in which it may sometimes be hard to extract particular data from systems. Business-related reasons include trade secrets, IP rights, and competitiveness. Legal reasons include the rights and freedoms of others, for instance, of other data subjects in their databases.

Altogether, these may be good reasons to not include inferred data, categories, and profiles too much in the scope of data subject rights. However, if data subject rights merely apply to personal data that they previously provided to data controllers, these rights are not very useful. Invoking these rights will not be very informative and valuable for data subjects, as it will not tell them anything new and it will not provide protection against decisions that data controllers may take upon analysis of these personal data.

If the focus is shifted from the textual interpretations of data subject rights and actual practices towards a more teleological and contextual interpretation of data subject rights, it is clear that data subject rights are intended to put data subjects more in control of their personal data.[118] In other words, with the support of data subject rights, data subjects are further empowered against and protected from practices by data controllers they may dislike. Following a more teleological interpretation of data subject rights, it could be argued that including inferred data, categories, and profiles are included in the scope of data subject rights makes sense: it provides data subjects with insights into what is done with their data and how any processing of their data leads to decisions on them and which decisions these are. These insights are needed for data subjects to exercise meaningful control over their data: only when they know what is happening with their data, data subjects can assess whether they agree to that or, if not, whether they want to invoke their control rights.

Clearly, this argument for teleological interpretations of data subjects is not novel. It is obvious that broader interpretations of data subject rights offer further empowerment and control for data subjects. However, our analysis adds a new argument to this: a narrow, mostly grammatical interpretation of data subject rights would give many data subject rights little or no meaning when applied to inferred data, categories, and profiles. It would not tell data subjects something new, something they did not already know.

Another argument, to the best of our knowledge not discussed in literature so far, is that our taxonomy of types of personal data offered in Section 2 covers all states of data processing. As shown in Fig. 1, data processing usually consists of different stages, but the data subject rights mostly seem to apply to the first stage (i.e., the input stage). If data subject rights (with a few exceptions mentioned in Table 1) do not apply to the further steps in the data processing, such as the data analysis ('throughput'), the analysis results ('output'), and the subsequent decisions on data subjects ('outcome'), they clearly do not cover the whole picture.

The argument that "data subject rights should *really* empower data subjects" strongly supports a broader interpretation of data subject rights that encompasses all stages in Fig. 1, rather than only the first stage of collecting personal data. The higher order information distilled from the 'raw' personal data, such as inferred data, categories, and profiles are not covered by grammatical interpretations of most data subject rights and therefore would not really empower data subjects. The information that data controllers generate is what is really important for data subjects to assess how they feel about the data processing. This is novel information to them and much more relevant than just the input data they already knew.

Beyond the need of being informed about the output of the data processing, there is also a pressing need for data subjects to take action regarding the data that is inferred on them and subsequently ascribed to them. Above we mentioned a number of examples (including that of a person being inaccurately labeled as a terrorist) which illustrate the importance of the ability of data subjects to delete, correct or stop personal data inferences.

When we argue for a broader interpretation of data subject rights, we do not argue for disregarding the data controller's legitimate interests. Article 23 of the GDPR applies regardless of how broad the scope of the data subject rights is. When justifiable, data controllers should be able to protect their interests in preventing (full) disclosures of inferences and details of their analytical processes. However, from a policy perspective, it seems beneficial to start from an interpretation that is more accommodating to those more vulnerable participants in the data economy.

5. Conclusion

In this article, we assessed to what extent data subject rights in the GDPR also apply to personal data generated by data controllers, such as inferred data and profiles. From a textual analysis of these rights, it becomes clear that most of these rights do not cover such categories of data, either because of the way some of these rights are phrased or because of the exceptions included limiting their practical value. Contrasting this textual analysis with a more teleological and contextual analysis it becomes clear that the textual interpretation does not offer the intended protection that data subject rights are supposed to offer. Sticking to the textual interpretation of data subject rights, which seems the most plausible from a legalistic perspective, means that these rights do not offer meaningful insights for data subjects in what is happening with their data. This means that data subject rights do not offer data subjects any significant empowerment against and protection from data controllers. Although this argument is not novel, our analysis shows that a broader interpretation of data subject rights would cover all stages of personal data processing, rather than only the initial stages. Furthermore, our analysis shows that a narrow interpretation (i.e., a grammatical rather than teleological) of data subject rights would give them little or no meaning when applied to inferred data, categories, and profiles. The data subject rights would simply not be informative for data subjects, it would not tell them anything new, anything they did not already know.

[117] Wachter, *supra* 50.

[118] See for instance Vrabec (2021), Solove, *supra* 99.

Real empowerment and protection could be achieved following a more teleological and contextual interpretation of data subject rights, in which personal data like inferred data, categories, and profiles are included in their scope. This is currently not the case when looking at case law, EDPB guidelines and opinions, and – perhaps most importantly – actual practices. Data controllers may have strong objections to this.[119]

Despite such objections, it seems important that the scope of data subject rights is interpreted more broadly to offer substantial protection and empowerment. Hence, it may be relevant to investigate how to overcome existing practical, business-related, and legal objections.

Practical objections concerning time, costs and efforts involved should not a priori outweigh data subject rights. These objections do not outweigh data subject rights for personal data obtained directly or indirectly from the data subject, so there seems no reason to strike a different balance for inferred data or profiles. Unless, perhaps, such data are much more difficult to extract. That is unlikely, because it is exactly the kind of information that the data analyses are supposed to deliver. If the data can be extracted for decision-making, they can also be extracted for data subjects invoking their rights.

Business-related objections concerning trade secrets, IP rights, and competitiveness can be addressed by aggregating or generalizing data, preventing reverse engineering of the underlying models. Providing highly technical information on how the data analytics works may not be meaningful information for the average data subject anyway. Providing descriptions of how the data are being processed and what exactly that means for a data subject is perhaps more useful and avoids digging too deep into trade secrets and IP rights. Data controllers may not want to reveal even descriptive information because it may reveal practices that data subjects dislike, but that would be insufficient reason to refuse compliance with data subject rights, as these rights are intended to counter this.

Legal objections include the rights and freedoms of others, for instance, of other data subjects in their databases. Inferred data, and profiles even more, may only be explainable in conjunction with personal data of many data subjects included in the models. To avoid disclosure of personal information of other data subjects, data protection authorities could play a role in this as trusted third parties. Instead of leaving it to individual data subjects to scrutinize these practices, data protection authorities could take this role. Data protection authorities already have the investigative powers and the required expertise to do this. However, they currently do not have the power to invoke data subject rights on behalf of the data subjects. Hence, for instance, if they discover that particular personal data is incorrect, they cannot request rectification on behalf of the data subject. The GDPR could be amended to make this work, obviously only with consent of the data subjects involved.

Declaration of competing interest

For this paper, there are no conflicts of interests to report.

Data availability

No data was used for the research described in the article.

[119] See for instance Bart Custers, Anne-Sophie Heijne, 'The right of access in automated decision-making: The scope of article 15(1)(h) GDPR in theory and practice', Computer Law & Security Review, Volume 46, 2022.