From knowing by name to targeting: the meaning of identification under the GDPR

Nadezhda Purtova*

Key Points

- Despite its core role in the EU system of data protection, the meaning of identification remains unclear in data protection law and scholarship while the spotlight focuses on the legally relevant *chance* of identification, ie identifiability.
- While Article 29 Working Party interpreted identification broadly, as distinguishing one in a group, this interpretation has been questioned in light of the CJEU decision in *Breyer*. This article tackles this uncertainty.
- This article offers an integrated socio-technical typology of identification where, in addition to the known identification types (look-up-, recognition-, session- and classification identification), targeting is added as a new identification type. To identify by way of targeting means to select a particular individual from a group as an object of attention or treatment in a single moment of time.
- The article clarifies the legal meaning of identification under the GDPR. It proposes a contextual interpretation of *Breyer*, which negates *Breyer's* restrictive potential and brings all identification types within the GDPR.
- The article concludes with a discussion of the implications of this reading of identification for data protection in terms the applicability of the GDPR to new data technologies and practices such as facial detection and non-tracking based targeted advertising, effects of certain privacy preserving technologies such as federated learning of cohorts, consequences for invoking data protection rights when identification is not possible, but also in terms of the need to clearly define the objectives of the data protection law.

Introduction

Identification, referring both to the process of identifying someone and the fact of being identified, is one of the boundary concepts of data protection law. It separates the data that is personal, i.e. relating to an identified or identifiable natural person, from non-personal, and thus triggers the applicability of the EU General Data Protection Regulation (the GDPR).[1] Yet, despite the high stakes attached to the meaning of this concept, relatively little attention is paid both in law and legal scholarship to what identification is. Therefore the chief issue tackled here is the meaning of identification under the GDPR.[2]

*Nadezhda Purtova, Faculty of Law, Economics, and Governance, Utrecht University, Utrecht, the Netherlands.
This contribution reports on the results of the project 'Understanding information for legal protection of people against information-induced harms' ('INFO-LEG'). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 716971). The article reflects only the author's view and the ERC is not responsible for any use that may be made of the information it contains. The funding source had no involvement in study design, in the collection, analysis and interpretation of data, in the writing of the report, and in the decision to submit the article for publication. I am especially grateful to Dr Michael Veale for his sharp comments and suggestions. I thank the journal editors Dr Jaap-Henk Hoopman, Dr Raphael Gellert, Prof. Ronald Leenes and the anonymous reviewer of this article for helping me sharpen this article's analysis.

1 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

2 The legal order of the Council of Europe, specifically Council of Europe Convention no 108 for the protection of individuals with regard to automatic processing of personal data of 28 January 1981, as updated in 2018 ('Convention 108+') also operates with the concept 'personal data' and defines it through the concept of identification as 'any information relating to an identified or identifiable individual' (Art 2(a) Convention 108+). Yet, examining the meaning of identification in the legal order of the Council of Europe is beyond scope of this article. It suffices to note that the European Court of Human Rights in its case law on Article 8 right to respect for private life referred to Convention 108+ and the definition of personal data, and recognized that '[s]uch data cover not only information directly identifying an individual . . ., such as surname and forename, . . . but also any element indirectly identifying a person such as a dynamic IP (Internet Protocol) address' (Registry of the European

The primary focus of the current scholarly attention lies on the adjacent concept of identifiability which refers to the *possibility* of identification, ie of being identified, in future.[3] This is not surprising since in practice whether or not a person is identifiable rather than identified is regarded as an easier criterion to meet and is therefore a *de facto* 'threshold condition' when determining the status of data as personal.[4] Some legal scholars discuss the meaning and legally relevant degree of identifiability,[5] pseudonymization, and true meaning and possibility of anonymization.[6] The debates among computer scientists tackle anonymization and reidentification techniques and their (in)effectiveness.[7] These discussions clarify the boundaries of application of data protection law and contribute to practical solutions for at least some of the data protection concerns, and as such are valuable and relevant. Yet, the meaning of identifiability is derived from and hence is secondary in relation to the primary concept of identification. Therefore any identifiability debate is at risk of being hollow when not underpinned with a robust understanding of identification. It makes little sense to argue if a natural person is 'identifiable' when it is not clear when a natural person would be 'identified' and what it means to identify somebody.

As the technologies to target a person evolve and test the boundaries of data protection, the meaning of identification becomes less clear, and the gap in understanding what it means to identify becomes increasingly more obvious and imperative to close.[8] A relatively recent case of such technological development is face detection and analysis used in 'smart' advertising boards.[9] Unlike with facial recognition where one's facial features are compared to pre-existing facial templates to establish if a person is known, face detection and analysis do not recognize people but 'detect' them and, in case of smart billboards, classify them into gender-, age-, emotion-, and other groups based on processing of their facial features to display tailored ads. The industry that develops, sells, and employs the technology argues that facial detection does not involve processing personal data,[10] eg because the chance of establishing who a person before the 'sensor' is close to null. In part this is due to the 'transient' nature of the processing, where raw data of an individual processed by the detection 'sensors' is discarded immediately.[11] The technology does not allow tracking a person and recognizing him or her over time either. To be clear, as will become apparent from further analysis, these industry arguments do not necessarily withstand legal scrutiny and it is highly likely that personal data will be processed in these contexts, if the proposed interpretation of identification is adopted. Yet, there is no uniform position on the interaction of face detection and data protection across the EU Member States.[12] For instance, the Dutch data protection authority considers face detection in the context of smart billboards as processing of personal data,[13] while its Irish and reportedly Bavarian counterparts are of the opposite view.[14] More similar debates and uncertainties are likely to emerge in other contexts where facial

Court of Human rights, *Guide to the Case-Law of the of the European Court of Human Rights. Data protection*, updated on 31 December 2021 <https://www.echr.coe.int/Pages/home.aspx?p=caselaw/analysis/guides&c> accessed 28 February 2022, analysis on page 7 and the cited case law). A brief study of the relevant case law suggests that the ECHR analysis also does not specifically address the meaning of identification as opposed to identifiability.

3 Article 29 Working Party 'Opinion 4/2007 on the concept of personal data' (WP 136, 20 June 2007), 12.

4 WP136, 12.

5 Frederic J Zuiderveen Borgesius, 'Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the New Data Protection Regulation' (2016) 32 Computer Law & Security Review 256; Paul Schwartz and Daniel Solove, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 NYU L Rev 1814, 1877.

6 Eg M Finck and F Pallas, 'They Who Must Not Be Identified—Distinguishing Personal From Non-Personal Data Under the GDPR' (2020) 10(1) IDPL 11.

7 Among most notable, Arvind Narayanan and Vitaly Shmatikov, 'Myths and Fallacies of "personally Identifiable Information"' (2010) 53(6) Communications of the ACM 24; Sweeney on *k*-anonymity (eg Latanya Sweeney, 'k-Anonymity: A Model for Protecting Privacy' (2002) 10(5) International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems 557) and responses to it, eg the works of Dwork and others on differential privacy, eg Cynthia Dwork and Aaron Roth, 'The Algorithmic Foundations of Differential Privacy' (2014) 9(3–4) Foundations and Trends in Theoretical Computer Science 211–407 <http://www.tau.ac.il/~saharon/BigData2015/privacybook.pdf> accessed 24 July 2020.

8 Nadezhda Purtova, 'The Law of Everything. Broad Concept of Personal Data and Future of EU Data Protection Law' (2018) 10(1) Law, Innovation, and Technology 40, 74; Peter Davis, 'Facial Detection and Smart Billboards: Analysing the "Identified" Criterion of Personal Data in the GDPR' (2020) 1 University of Oslo Faculty of Law Legal Studies Research Paper Series, <https://ssrn.com/abstract=3523109> accessed 27 July 2020.

9 Ibid.

10 Fraunhofer Institute for Integrated Circuits IIS, 'Emotion Recognition Software SHORE®: Fast, Reliable and Real-time Capable', <https://www.iis.fraunhofer.de/en/ff/sse/imaging-and-analysis/ils/tech/shore-facedetection.html> accessed 24 July 2020.

11 Damian George, Kento Reutimann and Aurelia Tamò-Larrieux, 'GDPR Bypass by Design? Transient Processing of Data under the GDPR' (2019) 9(4) International Data Privacy Law 285, 286.

12 As demonstrated by Davis (n 8).

13 Autoriteit Persoonsgegevens 'Normenkader Digitale Billboards' ('Normative Framework for Digital Bilboards') (25 June 2018) <https://autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/brief_branche_normkader_digitale_billboards.pdf> accessed 19 February 2021.

14 Data Protection Commissioner, 'Press Release on the Use of Facial Detection Technology in Advertising' (15 May 2017) <www.dataprotection.ie/docs/EN/15-05-2017-Statementon-use-of-Facial-Detection-Technology-in-Advertising/i/1634.htm> accessed 17 February 2018, no longer available; on hand with the author. The report published on 8 June 2017 by the Bavarian Data Protection Authority for the Private

analysis and sensing can be used, such as healthcare for pain or pulse detection, in the news sector for audience measurement, or in assisted driving,[15] video surveillance with face analytics,[16] but also online in the context of tracking-free advertising,[17] and in other cases of the 'transient' data processing. While the applicability of the GDPR would be the focus of debate in these contexts, the discussions will inevitably emerge also where the applicability of the GDPR is not in dispute, eg in the context of invoking data protection rights. Article 11(2) GDPR—under some caveats—exempts data controllers from complying with data subjects' data access and rectification requests, requests for erasure and restriction of processing, as well as data portability obligations where 'the controller is able to demonstrate that it is not in a position to identify the data subject'. The question will then be: what does it mean to identify? The definition of biometric data in Article 4(14) GDPR and pseudonymization in Article 4(5) GDPR also hinge on the meaning of identification.

To date, there have been disappointingly few attempts in the data protection legal scholarship, at least in English, at understanding identification beyond identifiability. In 2007 Leenes proposed a four-fold classification of identification. According to Leenes, there is more to identification than simply establishing one's civil identity, and we need to read identification broadly if we are to address the 'real privacy concerns'.[18] He distinguished look-up (l-), recognition (r-), classification (c-), and session (s-) identifiability.[19] A recent notable contribution to the debate on the meaning of identification is by Davis who examines the meaning of an 'identified natural person' specifically in the context of smart billboards and articulates the importance of looking into the meaning of 'identified' as a baseline for establishing the meaning of 'identifiable'.[20] However, Leenes, while examining the meaning of identification in data protection law, does so with a view to inform the information privacy debate across borders rather than to offer an interpretation of the specific legal concept of the EU data protection law, among others in light of the evolving case law of the Luxemburg Court, and Davis' analysis is limited to the legal status of data in the context of facial detection. Jasserand addressed the meaning of identification under the GDPR framework, but only when it concerns the definition of biometric data.[21]

In addition, there is a swirling stream of sociological and philosophical literature focusing on the related concepts of identity and anonymity. To name a few, in 1999 Gary Marx presented a sociological typology of what he called 'identity knowledge', which is the opposite of anonymity and hence I consider it equal to identification. He specified seven broad types of identity knowledge: legal name, locatability, pseudonyms linked to identity or location, pseudonyms that are not linked to name or location, pattern knowledge, social categorization, and symbols of eligibility/non-eligibility.[22] Helen Nissenbaum discussed the meaning and value of anonymity in the information age as 'unreachability'.[23] A range of scholars offer many accounts of the meaning and construction of identity, generally and in the context of ambient intelligence and profiling.[24] Against this backdrop the legal scholarly account of the meaning of identification is inadequate.

This lack of academic consideration might be partially explained by the fact that the Article 29 Working Party, an EU advisory authority on data protection under the former 1995 Data Protection Directive, defined what an identified person means in its 2007 opinion on the concept of personal data: '[i]n general terms, a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group'.[25] The same explanation arguably holds for the concept of personal data in the GDPR, since there are no fundamental differences between the definitions of personal data under

Sector (BayLDA). The report of the Bavarian data protection authority is not available online, but is referred to in Fraunhofer Institute for Integrated Circuits IIS, (n 10).

15 This is according to Fraunhofer Institute for Integrated Circuits IIS (n 10).

16 Eg *Bridges* case discussed further on in this article (*R (on the application of Edward Bridges) v The Chief Constable of South Wales Police and Secretary of State for the Home Department* [2019] EWHC 2341 (Admin) at 122-125 and *R (on the Application of Bridges) v South Wales Police* [2020] EWCA Civ 1058 at 46).

17 Eg Google's FLoC alternative to the tracking-based targeted advertising discussed further Chetna Bindra, 'Google Ads. Building a Privacy-first Future for Web Advertising' (25 January 2021, <https://blog.google/products/ads-commerce/2021-01-privacy-sandbox/>) accessed 19 February 2021.

18 R Leenes, 'Do They Know Me? Deconstructing Identifiability' (2008) 4(1&2) University of Ottawa Law & Technology Journal 135, 141–42. Although Leenes uses the word 'identifability', in effect he is talking about identification.

19 Ibid.

20 Davis (n 8).

21 Catherine Jasserand, 'Legal Nature of Biometric Data: From Generic Personal Data to Sensitive Data' (2016) 2 Eur Data Prot L Rev 297.

22 Gary T Marx, 'What's in a Name? Some Reflections on the Sociology of Anonymity' (1999) 15(2) The Information Society, 100.

23 Helen Nissenbaum, 'The Meaning of Anonymity in an Information Age' (1999) 15(2) The Information Society 141–44.

24 Eg contributions to Ian Kerr, Valerie Steeves and Carole Lucock (eds), *Lessons From the Identity Trail: Anonymity, Privacy and Identity in a Networked Society* (OUP, Oxford, New York 2009); Katja de Vries, 'Identity, Profiling Algorithms and a World of Ambient Intelligence' (2010) 12 Ethics and Information Technology 71–85.

25 WP136, 12.

the 1995 Directive and the Regulation. This approach includes identification by name, but also other modes of 'zoom[ing] in on a flesh and bone individual'.[26] The authority of the Working Party when it comes to the data protection on the ground is undoubted, and its opinion on the concept of personal data is the most comprehensive and influential guideline for the controllers as to how this concept should be used in practice. The general perception of the meaning of identification under the GDPR following from the WP29 interpretation is thus that it is broad, flexible, and generously accommodating to the realities and challenges of the modern data processing practices.[27] Indeed, the meaning of identification as distinguishing a person from a group should bring the cases of targeted advertising, profiling, and others where the name of a person is of no consequence to the protective bosom of the GDPR. Perhaps for this reason the data protection scholarship seems to be comfortably content with the status quo in law and literature.

However, the status quo has been resting on shaky grounds. The position of the Working Party, and hence the 'distinguished from' approach to identification, are not formally binding. The Court of Justice of the European Union (CJEU), the only body with authority to issue binding interpretations of the GDPR, was long silent on the meaning of identification. While the Court did follow the Working Party in interpreting the 'information' and 'relating to' elements of the concept of personal data in *Nowak*,[28] it also has a record of not following the lines of interpretation chosen by the WP29 earlier.[29] To complicate matters further, the Court in its 2016 *Breyer* decision[30] appeared to have invalidated the understanding of identification as distinguishing or being distinguished from a group, advanced by the Working Party and granting the GDPR protection a broad reach. Without any detailed consideration about the meaning of identification, the Court in *Breyer* dismissed a dynamic IP (Internet Protocol) address as an identifier sufficient to identify a person,[31] while one of the core functions of an IP address is exactly to distinguish one web visitor, or at least a location on the network, from another.[32]

This brief consideration seems to restrict the interpretation of identification under the GDPR to the identification by name or a similar unique identifier representing one's civil identity, the narrowest meaning of identification possible.[33] This effectively takes cookies, IP addresses, and other online trackers,[34] and with them a large part of online tracking and discrimination, but also not name-tied individual profiling and (real-time) automated decision-making, among others enabled through some of the new technologies such as facial detection, outside of the scope of the data protection law, and deprives people affected by these practices of legal protection that the GDPR would have granted, was the identification interpreted broadly. The very limited scholarly commentary on the *Breyer* case has largely overlooked this remarkable and consequential departure of the CJEU from the WP29 interpretation.[35] Hence, the question remains: how should identification under the GDPR be understood?

This article will answer this question in two steps. First, it will examine the meaning of identification outside of the legal context (the Section 'Meaning and Socio-Technical Approaches to Identification outside of the GDPR'). It will offer an integrated typology of identification as a process and result of distinguishing a person in a group. The typology builds on three prominent socio-technical accounts of identification: four identifiability types by Leenes, seven types of identity knowledge by Marx, and anonymity as unreachability by Nissenbaum. In addition to the established types, I will identify targeting as a new identification type, where to identify by way of targeting means to select a particular individual from a group as an object of attention or treatment in a single moment of time. The argument will build, among others, on the literatures on calculated publics, profiling in recommender systems, price, and content personalization. Second, I will focus on the legal meaning of identification under the GDPR. I will build a case that all five identification types not limited

26 Ibid 13–14.

27 See eg Lee A Bygrave and Luca Tosoni, 'Article 4(1) Personal data' in Christopher Kuner and others (eds), *The EU General Data Protection Regulation (GDPR). A Commentary* (OUP, Oxford 2020).

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

29 A recent example is a decision in *Google Spain SL, Google Inc v Agencia Española de Protección de Datos and Mario Costeja González* Case C-131/12 [2014] ECLI:EU:C:2014:317 [31] et seq. where the Court found a search engine provider a controller, contrary to the earlier position of the Article 29 Working Party.

30 *Patrick Breyer v Bundesrepublik Deutschland*, Case C-582/14, [2016] ECLI:EU:C:2016:779.

31 *Breyer* [38].

32 Davis (n 8) 17 et seq.

33 Ibid 17.

34 Except a limited number of cases when the data processed also contains information revealing identity, eg vanity searches.

35 The author was able to locate very few papers published by the time of writing that discuss *Breyer* and none of them, besides Davis, discuss the Court's stance on the meaning of 'identified' in any significant detail. The papers reviewed include Frederic J Zuiderveen Borgesius, 'Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition' (2017) 3(1) European Data Protection Law Review 130; Alan Reid, 'The European Court of Justice Case of Breyer' (2017) 1 (2) Journal of Information Rights, Policy and Practice; Bygrave and Tosoni (n 27).

to civil identity identification are covered by the GDPR meaning of identification. It is an easy conclusion to draw if one follows a non-binding interpretation of Article 29 Working Party that to identify means to distinguish one in a group. This approach will be detailed in the section 'The Article 29 Working Party Interpretation of the GDPR'. In the section 'Meaning of Identification in CJEU's case law' I review the CJEU case law with relevance to the meaning of identification, including *Breyer* and its potentially restrictive impact. I then propose a contextual interpretation of *Breyer* in light of the facts of the case, which negates *Breyer's* restrictive potential and brings all types of identification, including non-civil identity ones, within the meaning of identification under the GDPR. The section' Conclusion: What This Means for Data Protection' will conclude with a discussion of the implications of this broad reading of identification for EU data protection law practice and research.

Meaning and socio-technical approaches to identification outside of the GDPR

Non-legal, or ordinary meaning of concepts always provides a foundation of their use in law, sometimes adjusted to the legislative history and intent, objectives and general system of the piece of legislation at hand. In English, identification means 'the action or process of identifying someone or something or the fact of being identified'[36] and to identify means 'to establish or indicate who or what (someone or something) is . . .; recognize or distinguish . . . ,'[37] where 'to distinguish' refers to recognition or treating of someone or something differently.[38] The verb 'to individuate' is a synonym of 'to distinguish from others' and 'to single out'.[39] According to Davis, the linguistic equivalents chosen in at least 14 non-English official EU language versions of the GDPR have a similar meaning.[40] Consequently, a person is identified when it is established who he or she is, when he or she is recognized from being known before or from some characteristics, or when that person is recognized as a distinct individual or treated differently. However, in addition to the dictionary meaning, there are various sociological and socio-technical analyses of what identification is. Without aiming at comprehensive cataloguing of these analyses, the remainder of this section will consider three prominent accounts of the meaning of identification: the four types of identification by Leenes, the seven types of identity knowledge by Marx, and the account of anonymity as unreachability (and hence identification as reachability) by Nissenbaum.

Operational definitions of identification: Leenes and Marx

Leenes and Marx propose what can be considered operational definitions of identification, i.e. they list practices that—when present—indicate that identification is taking place. Leenes relies on the conceptualization of identification as the process or fact of being singled out or 'individualized within a set of subjects, the identifiability set'[41] and distinguishes four types of identification:[42] look-up (l-), recognition (r-), classification (c-), and session (s-) identification.

1. The *look-up (l-) identification* is an identification of a named individual by an identifier, such as a name, telephone or passport number, and even an IP address, when there is a registry, directory, or a table that connects that identifier to a named individual (ie his/her civil identity). Using an l-identifier, an individual can be 'looked up' in the real world, hence the name.[43]
2. *Recognition (r-) identification* refers to the identification of an individual without a reference to his/her civil identity and requires presence or activity of an individual. An individual is identified from being known before or by presenting certain features, ie 'she presents an identifier, token or feature set (e.g. description of physical appearance), known or recognizable as valid by the recipient, to the entity performing the identification'.[44] For instance, a token (eg a cloak room token) allows the recipient (a cloak room clerk) to recognize the holder as someone, or something, or as being entitled to something (eg to receive a coat checked in in the cloak room).[45] Facial recognition is an example of r-identification. An individual's face is compared to a facial template made during a preceding interaction with that individual, to verify if that individual is, eg a repeated visitor of a store, or has authorization to enter a

36 A Stevenson, J Pearsall and P Hanks (eds), *Oxford Dictionary of English* (3rd edn, OUP, Oxford 2010) 868.

37 Ibid 869.

38 Ibid 509.

39 Ibid 891.

40 Davis ((n 8) 18) considered 15 out of the 23 non-English versions.

41 Leenes (n 18) 147–48.

42 Leenes calls them types of identifiability, but the types he proceeds to describe do not refer to the possibility of identification in future but rather to the process of identification. Therefore I consider the typology he proposes to be a typology of identification.

43 Leenes (n 18) 148.

44 Ibid 150.

45 Ibid.

building (if facial recognition is used as a method of biometric authentication). Without the need to establish an individual's civil identity, r-identifiers connect several interactions with one individual together and 'enable personalisation of experience'.[46] Persistent cookies, device-generated advertising IDs, and IP addresses are examples of r-identifiers, and ecommerce is one area where establishing one's civil identity is not necessary and r-identification is used a lot,[47] among others for consumer profiling and targeted advertising.

3. In case of the *classification (c-) identification*, individuals are 'identified as members of a particular [preexisting] group of category'.[48] The purpose is not to establish an individual's civil identity or recognize him or her, but to classify an individual as a member of one or several groups. While categorization is often achieved through observing individuals over time, eg through (online) tracking and use of l- or r-identifiers, it can also exist independently. In this case, a preexisting knowledge of the categories and of the attributes that put an individual in one or more of these categories is required.[49] Facial detection technology allowing to segment passers-by of smart ad boards into audience segments and demonstrate segment-tailored ads would be an example of classification not relying on tracking and l- or r-identification.
4. *Session (s-) identification* aims to track an individual during a particular interaction, and the lifetime of the s-identifiers is restricted to the duration of that interaction.[50] An example is session cookies that allow an online shop to individualize a visitor's shopping experience, eg make sure the website remembers the items in a shopping basket.

Marx presents a sociological typology of identification (the opposite of anonymity) that he understands as 'identity knowledge'.[51] According to Marx, there are at least seven types of identity knowledge, also reflecting degrees of identifiability: (i) legal name, (ii) locatability, (iii) pseudonyms that can be linked to legal name and/or locatability, (iv) pseudonyms that cannot be linked to other forms of identity knowledge, (v) pattern knowledge, (vi) social categorization, and (vii) symbols of eligibility/noneligibility.[52]

1. Identification by a 'legal name' involves a full name that is presumed unique in a given context (eg only one child named John Smith is born to a particular set of parents) and connects to the information 'biological or social lineage' and a large amount of other information about a person.[53]
2. Identification as 'locatability' involves 'reachability' of an individual by an address, actual or in the cyberspace (an IP address would be a good example of reachability in the cyberspace). While it does not require knowledge of an individual's civil identity or a pseudonym, it does imply the ability to reach a person and treat him or her in a certain way, eg block or grant access, charge or penalize.[54]
3. Identification by 'pseudonyms that can be linked to legal name and/or locatability' involves 'alphabetic or numerical symbols', ie pseudonyms, that link to the person's name or address. Such identification usually involves a third trusted party which serves as a buffer to facilitate a compromise between preserving one's real identity or address but achieving some degree of identification.[55]
4. Identification by means of 'pseudonyms that cannot be linked to other forms of identity knowledge' refers to the identification by symbols, names or pseudonyms that, 'under the normal circumstances', cannot be connected to a person, either due to special anonymization measures or due to the fact that the identifier is fraudulent, such as the pseydonyms used by spies or con artists.[56]
5. Identification by 'pattern knowledge' involves identification by reference to a repeated observation of 'distinctive appearance or behavior patterns'[57], not connected to the name (civil identity) or the locatability of a person. Examples that Marx cites are recognizing someone you repeatedly met on the metro as someone you 'know', recognizing a donor by a repeated pattern of donation, a criminal by a pattern of his crimes, etc.[58]
6. Identification may happen by 'social categorisation' since 'many sources of identity are social'. [59] Hence, individuals can be identified by gender, ethnicity, organizational membership and other classifications

46 Ibid.
47 Ibid.
48 Ibid 151.
49 Ibid.
50 Ibid 152.
51 Marx (n 22) 100.
52 Ibid 100.
53 Ibid.
54 Ibid 101.
55 Ibid.
56 Ibid.
57 Ibid.
58 Ibid.
59 Ibid.

which do not 'differentiate the individual from others sharing them'.[60]

7. Identification by 'symbols of eligibility/noneligibility' involves 'certification' where the possession of knowledge such as possession of a code word, artifacts, such as a ticket or a smart card, or a skill, eg ability to swim, warrants a particular treatment, eg entitlement for a reimbursement, or a sanction for system abusers.[61]

Nissenbaum's anonymity as unreachability

While Helen Nissenbaum does not discuss the meaning of identification directly, her work on the meaning and value of anonymity is of immediate relevance for conceptualization of identification. Identification and anonymity are the opposites, and therefore the meanings of these concepts are intimately related. In 'The Meaning of Anonymity in an Information Age'[62] Nissenbaum argued that the value of anonymity has traditionally been to ensure unreachability, ie to ensure that when one acts in a certain way, no one would knock on his door 'demanding explanations, apologies, answerability, punishment or payment'.[63] While the best way to ensure this in the past was to remain nameless, 'the power of information technology to extract or infer identity from non-identifying signs or information' has changed this, and remaining nameless or withholding other unique persistent identifiers in place of a name such as a social security or a passport number is no longer sufficient to protect unreachability.[64] The current dangers of data processing are not limited to eg one government body connecting its record on someone to the record of that person with another government body via, eg a social security number.[65] As the advertising industry puts it, '[t]he beauty of what we do is we don't know who you are … We don't want to know anybody's name. We don't want to know anything recognizable about them. All we want to do is … have these attributes associated with them'.[66] This analysis suggests that, if anonymity should be understood as unreachability, identification should be understood as the process or the fact of someone being reached.

Synthesis: towards an integrated operationalization of identification and targeting as identification

The remainder of this section is a proposal for an integrated socio-technical operationalization of identification. The question to be answered is: which practices constitute identification in concrete terms. The two typologies reviewed may already be considered operationalizing identification. This section will integrate and refine them, also taking into account Nissenbaum's perspective on identification as reaching a person, to reflect the conceptual meaning of identification fully. The exercise identifies a new type of identification not articulated before, ie targeting.

Juxtaposing identification typologies

The three perspectives on identification reviewed above agree with each other well, and reflect an understanding of identification in line with its dictionary meaning, as a process or fact of recognizing or distinguishing someone. The typology by Leenes is better suited to be a foundation of an integrated understanding of identification compared to that by Marx. The typology offered by Marx is less suitable as a typology of 'identification', because, as Marx states, it reflects not only the meaning but the degrees of identifiability as a 'possibility of' identification rather than purely of identification as a fact or a process. As a result, it is not entirely clear where Marx draws a line between identification and identifiability, eg if he considers identification by name as the only mode of true identification (type 1) while the remaining types are meant to refer to degrees of identifiability. Leenes' typology is more clear-cut, i.e. focuses on the essential features of each type without overlap, but also provides a nuance that Marx's typology does not have. To name one example, Marx does not account for the lifespan of identifiers as Leenes does.[67] All the types identified by Marx fit under one and some under two of the types distinguished by Leenes (as presented in Table 1).

Three out of seven types of identification distinguished by Marx match Leenes' identification types: Marx's identification by a legal name (type 1) fits within Leenes' look-up identification, identification by pattern

60 Ibid.

61 Ibid.

62 Nissenbaum (n 23) 141–44.

63 Ibid 142. See also Daniel Solove, *Understanding Privacy* (Harvard University Press, Harvard 2008) 125 where Solove expresses a similar view on identification, namely, that identification *links* the digital person created by aggregation of data points to a person in real space.

64 Nissenbaum (n 23) 142.

65 Solon Barocas and Helen Nissenbaum, 'Big Data's End Run around Anonymity and Consent' in Julia Lane and others (eds), *Privacy, Big Data, and the Public Good. Frameworks for Engagement* (CUP, New York 2019) 44–75.

66 Cindy Waxer, 'Big Data Blues: The Dangers of Data Mining' 4 November 2013 Computerworld, <http://www.computerworld.com/s/article/print/9243719/Big_data_blues_The_dangers_of_data_mining> cited in Barocas and Nissenbaum (n 65) 54.

67 As evident in case of session identification.

Table 1. Relationship between Leenes' and Marx's typologies

	L-IDENTIFICATION		R-IDENTIFICATION		C-IDENTIFICATION	S-IDENTIFICATION
Leenes	*Establishing civil identity via a register that links an identifier (name, passport number, etc.) to a person in the real world*		*Recognizing a person as 'known' or eligible by 'an identifier, a token or a feature set' seen as valid.*		*Classification of a person as a member of a pre-existing group or category.*	*Tracking of a user during one interaction, where the lifetime of an identifier is limited to a session.*
Marx	1. Identification by a legal name;	3. Pseudonyms that can be linked to legal name and / or locatablity; 2. Locatability;	4. Pseudonyms that cannot be linked to a legal name or locatability; 5. Pattern knowledge;	7. Identification by symbols of eligibility	6. Social categorisation;	4. Pseudonyms that cannot be linked to a legal name or locatability;

knowledge (type 5) fits under recognition identification; social categorization (type 6) is equivalent to Leenes' classification identification. Yet, four out of seven types display characteristics of two identification types according to Leenes: identification by a pseudonym that can be linked to civil identity (type 3) and locatability (type 2) fit both under the look-up and recognition identification. The former is the case because Marx presumes a pseudonym to be connected to a person's civil identity and only separated from that identity by a third trusted party. This renders that person identifiable in the look-up sense of identification. Similarly, a physical or cyber (eg IP) address may serve as a look-up identifier when connected to a person's real world identity via a registry, like it is the case with static IP addresses. At the same time, both the pseudonym and locatability identifiers can serve as recognition identifiers, eg to recognize and track interaction with individuals over time, where, albeit possible, the establishment of who a person in the real world is not necessary, eg for the targeted advertisement purposes. Marx's identification by symbols of eligibility (type 7) fits both under recognition- and classification identification. If the distinctive feature of this mode of identification is the resulting eligibility for a particular treatment, it can be achieved both by using r- and c-identifiers: r-identifiers when certain treatment is triggered by a token or another identifier tagging a person as 'known' or 'eligible', eg to enter a building based on facial recognition, and c-identifiers when the treatment is triggered by a person displaying characteristics of a group: male or female, reader of detective novels, at high risk of diabetes, etc. Identification by pseudonyms not linkable to civil identity or locatability (type 4) fits both under recognition- and session-identification, depending on the lifetime of the identifier.

Targeting—new identification type

Considering the two typologies together and in light of Nissenbaum's work reveals another mode of identification that neither Leenes nor Marx articulate as a distinct type. Yet, this identification mode is implied in the understanding of identification as individuation and distinguishing one from a group, including reaching a particular person, and has sufficient defining features to be distinguished as a separate identification type. This is targeting (or t-identification).

To identify by way of targeting means to select a particular individual from a group as an object of attention or treatment in a single moment of time. *T-identification* is the most basic mode of individuation. It does not aim at establishing civil identity. Unlike recognition- and to

some degree session identification, t-identification does not rely on a persistent identifier such as an IP address or cookies and does not aim to recognize an individual during a future encounter. Instead, targeting occurs in real time and at the single moment of contact. Unlike classification identification, targeting does not aim to identify an individual as a member of one or several groups and does not require a pre-existing knowledge of the categories and of the attributes that put an individual in these categories. Instead, the purpose of targeting is pure individuation, zooming in on a particular individual who is distinct from others. This can be done in order to subject that individual to tailored treatment or content.

Targeting can be achieved either by means of a unique identifier that does not need to be persistent, or can be based on the rich dataset, eg provided by a device in real time and allowing unique characterization of an individual that distinguishes that individual from others.

An example of t-identification by means of a unique identifier is identification by a media access control (MAC) address. A MAC address is a unique identifier usually assigned to a device by its manufacturer as a hardware address for communication in a network. A persistent MAC address enables continuous monitoring of movements of a particular mobile device which would amount to session identification, or recognition of the same device on a repeated encounter which constitutes recognition identification. As a countermeasure against such tracking, many device manufacturers introduced randomization of MAC addresses while devices are scanning for networks. However, even when randomized MAC addresses do not allow for tracking devices across time, each random MAC address—although short-lived—is still unique and distinguishes one unique device from another for the purposes of communication in a network. T-identification is the use of a MAC address, whether persistent or randomized, solely in order to distinguish one device (and its user) from another 'in a single moment of time' rather than facilitate tracking or recognition.

A human face is another unique identifier that can be used for recognition of individuals (in the sense of r-identification) when their facial data is matched to pre-existing facial templates of known individuals. However, facial data does not always have to be matched to facial templates and does not have to lead to recognition. In t-identification, facial data is used to distinguish one unique face from another in a single moment of time which is not aimed at recognition or tracking. This can be done in the context of crowd management, or in order to infer demographic and emotional data from facial features and display advertising tailored accordingly.

An individual can be t-identified on the basis of a unique characterization on the basis of a rich dataset. When discussing identification by pattern knowledge, Marx writes:

> Some information is always evident in face-to-face interaction, because we are all ambulatory autobiographies continuously and unavoidably emitting data for others' senses and machines. … This has been greatly expanded by new technologies.[68]

With t-identification based on a rich dataset, the unique identifier is not unique facial features or a MAC address, but that unique data-driven 'autobiography' that a user's machine is broadcasting in real time and that uniquely distinguishes that machine and its users from others. The difference with the t-identification based on a single identifier is that—in addition to purely distinguishing an individual—the rich dataset also provides his or her description, a unique characterization.

Browser fingerprinting is one instance of such information-emitting autobiography. Browser (or web-) fingerprinting is a method used to collect detailed information about the machine of a website visitor, including a browser type and version, operating system, language and security settings, screen resolution, and other parameters. These data form a 'fingerprint' that can be used to recognize browser users when they are encountered again. But recognition is not the only use of browser fingerprinting. The data captured in the 'fingerprint' can be sufficiently rich to enable a unique characterization of a person, to distinguish a person without a reference to earlier encounters. This would constitute identification in the sense of targeting.

Sparse or dense matrices and embedding are examples of techniques that can be used here. To be t-identified on the basis of a rich dataset, one is characterized or mapped in relation to a multiplicity of dimensions or axes within a multidimensional space, where an axe can be attributes, such as facial and physical dimensions, interests, behaviour, or the attributes of the surrounding context, eg a device, as a container for such behaviour. Seaver suggests that the spread of sensory technology is 'expected to provide even more contextual signals', such as the ambient noise level, acceleration, etc.[69]

T-identification based on rich datasets shares some similarities with classification identification in the sense

68 Marx (n 22) 101.

69 Nick Seaver, 'The Nice Thing about Context is That Everyone Has It' (2015) 37 (7) Media, Culture & Society 1102.

that the unique characterization is done based on static or dynamic characteristics or attributes that could belong to categories or groups. The difference is that, unlike with classification which is essentially a result of putting people in one or several boxes populated by many (Russian, Dutch, 25-year-olds, blond or dark haired), in case of t-identification, using the vocabulary of differential privacy, $k = 1$.[70] In other words, the more axes or parameters of characterization are used, the fewer people share the same location on the axes. The more parameters are included, the closer the characterization is approaching unique. Moor and Lury call this 'personality construction' which is based on fragments of a personality of an individual relevant to some actors in some contexts,[71] what Delueze called dividuals.[72] Unlike with classification and categorization, targeting is not concerned with a person as a member of one or several groups, but aims at personalization. An individual is characterized by an overlap of a very large number of attributes and classifications, where any group attributes and classifications are increasingly not along the static and socially constructed socio-demographic lines, but are algorithmically constructed. The resulting overlap is relatively unique.

T-identification on the basis of rich datasets is to a large degree a product of a shift of classification practices towards algorithmic classification. As a result, the categories in which people can be classified become less stable and obvious and less transparent to a person being categorized or even to those doing the categorization. While categorization based on widely used and known and relatively static parameters such as age, social status, or ethnic origin and other socio-demographic criteria is more obvious and transparent, categorizations are increasingly done in the form of the so-called 'calculated publics"'[73] where the categories and attributes are not socially but algorithmically constructed. As a result, the categories are dynamic, interactive, iterative, descriptive but increasingly more generative,[74] and so less obvious and transparent.

At present, identification by persistent identifiers in the sense of l-identification, recognition or session identification is certainly more common and more known. Yet, t-identification might quickly become more prevalent as a result of an interplay of a number of developments. The first such development is a push towards more personalized content,[75] advertising and pricing.[76] Second, t-identification, especially based on rich datasets, will likely become more widespread as a part of a larger move towards context-aware computing,[77] According to Seaver, 'we are in for a future where data mining concerns itself increasingly with the determination of context, drawing on a range of signals to personalize more precisely than the unified "person"'.[78] Finally, the popular perception of data processing risks is connected to names and other persistent identifiers, and the focus of the 'privacy-preserving technologies' and enforcement efforts also lies on persistent look-up and recognition identifiers. Targeting identification enables to reach a unique individual with tailored content or treatment without relying on those identifiers tainted by public and enforcement attention, and therefore may well become the winning strategy growing in popularity.

Temporal dimension of identification

Looking at the resulting types of identification from the perspective of Nissenbaum's work, it becomes clear that the various identification types can also be characterized based on a temporal dimension. Only the look-up and recognition identification types involve a (somewhat) persistent identifier that allows distinguishing, or reaching, a particular person through time. Indeed, only look-up and recognition identification, eg by name, address, a static IP address, a token or a repeating pattern of behaviour enable longitudinal observation, holding a person accountable for his or her past actions, sanctioning, holding eligible or rewarding a person based on something that took place in the more or less remote past. Session identification also has this temporal feature, albeit limited to the lifetime of one interaction, eg a website 'remembers' which item the visitor put in the basket. Classification and targeting identification clearly do not have such a longitudinal element. There an identified person is 'reached' or distinguished based on the

70 *k* refers to the number of people fitting into a group or category.

71 Liz Moor and Celia Lury, 'Price and the Person: Markets, Discrimination, and Personhood' (2018) 11(6) Journal of Cultural Economy 501–13.

72 Gilles Deleuze, 'Postscript on the Societies of Control' (1992) 59 October 3–7.

73 Tarleton Gillespie, 'The relevance of algorithms' in Tarleton Gilespie, Pablo Boczkowski and Kirsten Foot (eds), *Media Technologies: Essays on Communication, Materiality, and Society* (Cambridge, MA: MIT Press) 177. See also Moor and Lury (n 71).

74 Eg as Seaver observes, '[i]n demographic marketing, groups of people and groups of products are mutually defining: brand strategists understand pizzas in terms of people and people in terms of pizzas'. (Nick Seaver, 'Algorithmic Recommendations and Synaptic Functions' (2012) 2 Limn <https://limn.it/articles/algorithmic-recommendations-and-synaptic-functions/> accessed 19 February 2021).

75 Nick Couldry and Joseph Turow, 'Advertising, Big data, and the Clearance of the Public Realm: Marketers' New Approaches to the Content Subsidy' (2014) 8 International Journal of Communication 1710.

76 As illustrated in Moor and Lury (n 71).

77 Seaver (n 69) 1101–09.

78 Ibid 1103.

Table 2. Integrated typology of identification

L-IDENTIFICATION	R-IDENTIFICATION	S-IDENTIFICATION	C-IDENTIFICATION	T-IDENTIFICATION
Establishing civil identity via a register that links an identifier (name, passport number, etc.) to a person in the real world	Recognizing a person as 'known' or eligible by 'an identifier, a token or a feature set' seen as valid.	Tracking of a user during one interaction, where the lifetime of an identifier is limited to a session.	Classification of a person as a member of a pre-existing group or category.	Selecting a particular individual from a group as an object of attention or treatment in a single moment of time.
Temporal dimension				
Persistent identifiers allow longitudinal tracking		Limited persistent identifiers; Longitudinal tracking limited to duration of one session	No persistent identifiers No longitudinal tracking	

features he or she presents in real time, with real time consequences. What results is an integrated typology of identification as presented in Table 2.

The Article 29 Working Party interpretation of the GDPR

Broad meaning of identification

The GDPR does not directly define identification. The only relevant provision of the GDPR is the Article 4(1) definition of personal data.

> 'Personal data is any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.'

The definition refers to an 'identified and identifiable natural person', explaining that 'an identifiable natural person is one who can be identified'. Recital 30[79] names non-name identifiers such as RFID that can enable identification, but is inconclusive as to whether or not an individual is 'identified' by a non-name identifier or only 'identifiable'. Recital 26 provides some guidance on when a natural person should be considered identifiable and establishes the test of 'the means reasonably likely to be used . . . to identify', calling for all objective factors of the case to be considered. Yet, while 'identifiable' in the definition clearly refers to the possibility of identification, ie of being identified, no explanation of what 'identified' means is given.

Some explanation is provided by the Article 29 Working Party. The Article 29 Working Party, an EU advisory authority on the matters of data protection under the 1995 Data Protection Directive (the DPD), adopted a non-binding opinion on the concept of personal data (WP136).[80] The current status of the opinion is not certain. On the one hand, it concerns the concept of personal data in the old DPD and not the GDPR, the Article 29 Working Party itself no longer exists and is substituted by a new advisory authority—the European Data Protection Board (EDPB). Shortly after coming to existence, this functional equivalent of the Article 29 Working Party endorsed a number of Article 29 Working Party opinions, yet WP136 is not among these.[81] On the other hand, an argument can be made that the opinion retained its significance also under the GDPR, since the concept of personal data has not undergone significant changes.[82] While in future the EDPB may choose to issue its own

79 Recital 30 GDPR reads: 'Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.'

80 WP 136 (n 3).

81 See <https://edpb.europa.eu/our-work-tools/general-guidance/endorsed-wp29-guidelines_en> accessed 13 June 2022.

82 See eg Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECLI:EU:C:2017:994, Opinion of Advocate General Kokott [3].

GDPR-specific guidelines on the concept of personal data and take a different view on what identification means, it has not done so yet and its work programme for 2021–22 has given priority to other key data protection concepts such as legitimate interest.[83] For this reason, the Article 29 Working Party opinion remains influential and will be considered as such here.

WP29 adopts an understanding of identification which is in line with its dictionary meaning: '[i]n general terms, a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group', and 'the natural person is "identifiable" when, although the person has not been identified yet, it is possible to do it'.[84] Throughout the text of the opinion the WP29 also uses other formulations as stand-ins for 'to distinguish from the group': to 'single out a particular person' or to 'zoom in on a flesh and bone individual'.[85]

The WP29 explains that identification is achieved through the so-called identifiers. The identifiers are 'particular pieces of information . . . which hold a particularly privileged and close relationship with the particular individual',[86] like a name, 'outward signs of the appearance of this person, like height, hair colour, clothing, etc. . . or a quality of the person which cannot be immediately perceived, like a profession, a function'.[87] The WP29 does not shed any light on the criteria that determine that close or privileged relationship. Intuitively, not all of the examples of identifiers hold that special and privileged position in relation to an individual. While a name, a social security number, and perhaps some appearance traits, eg a face, can be said to be in that particular relationship to an individual due to a psychological bond (eg with a name and a face) or because they are unique to that individual (eg a face and a social security number), it is difficult to call a relationship between an individual and his or her hair colour, height or profession 'privileged' or particularly close, since hundreds of thousands of people may share these characteristics, and individual can change at least some of those characteristics (eg by dying the hair, wearing hilled shoes or changing a career). Which aspect of the relationship between a piece of information and an individual makes it special, making that piece of information an identifier according to the WP29, remains guesswork. Therefore a simpler and more consistent way to define an identifier would be as a piece of information that, alone or in combination with other identifiers, distinguishes a person in a group.

What requires more attention though is what the WP29 understands as 'direct' and 'indirect' identification, and consequently when an individual is identified (or identifiable) 'directly' and 'indirectly'. The WP29 explains that a person may be identified or identifiable either directly or indirectly.[88] In other words, 'directly or indirectly' in the definition of personal data ('an identifiable natural person is one who can be identified, directly or indirectly') applies to an identified as well as identifiable natural person, and not just to the latter. This follows from the legislative history of the definition of personal data in the 1995 Directive where the commentaries to the amended Commission proposal also distinguish two ways in which a person may be identified: '[A] person may be identified directly by name or indirectly'.[89]

Some confusion may occur where the commentary the WP29 cites explains that

> 'a person may be identified directly by name or indirectly by a telephone number, a car registration number, a social security number, a passport number or by a combination of significant criteria which allows him to be recognized by narrowing down the group to which he belongs (age, occupation, place of residence, etc.).'[90]

The commentary suggests that only identification by name should be considered as direct identification, and identification by other single identifiers such as a telephone number, a car registration number, a social security number, or a passport number should be considered indirect, just as the identification by a combination of significant criteria that narrow down the group to which a person belongs should be considered as a case of indirect identification. While the WP29 does not explicitly disagree with this understanding, its further explanation testifies to this effect. While the opinion is quite detailed, it is not always conclusive for the purposes of our analysis, specifically, because the explanation is structured along the lines of 'directly' versus 'indirectly' identified or

83 European Data Protection Board, 'EDPB Work Programme 2021/2022' (available online at <https://edpb.europa.eu/about-edpb/about-edpb/strategy-work-programme_en>, accessed 30 may 2022).

84 WP136 12. The 'distinguished from the group' understanding of identification seems to have been broadly adopted in Europe. The European Agency for Fundamental Rights and the Council of Europe explain that identification 'requires elements which describe a person in such a way that he or she is distinguishable from all other persons and recognizable as an individual'. *Handbook on European data protection law* (European Agency for Fundamental Rights and Council of Europe, 2018) <https://www.echr.coe.int/Documents/Handbook_data_protection_ENG.pdf> 89.

85 WP136 13–14.

86 Ibid 12.

87 Ibid 12.

88 Ibid 12.

89 Ibid 12–13.

90 Ibid.

identifiable, rather than 'identified' and 'identifiable'. The result is that it is not always possible to separate the WP29 considerations that concern 'identified' from the considerations concerning 'identifiable'. For this reason, this analysis is bound to be an interpretation of the WP29 opinion, rather than its restatement.

Regarding 'directly' identified or identifiable persons, the WP29 observes that the name 'is indeed the most common identifier, and, in practice, the notion of 'identified person' implies most often a reference to the person's name'.[91] Yet, 'a name may itself not be necessary in all cases to identify an individual'.[92] Other 'unique identifiers' can be used to distinguish one person from another, such as identifiers assigned to persons in computer files (eg file numbers), or web traffic surveillance tools,[93] presumably, such as cookies, IP addresses, and other online identifiers. Since a computer is the individual's contact point, the ability to identify an individual 'no longer . . . requires the disclosure of his or her identity in the narrow sense', and does not 'necessarily mean the ability to find out his or her name'.[94] Perhaps, following the same logic, the final definition of personal data in the 1995 Directive which transitioned into the GDPR without significant changes does not follow the Commission verbatim and lists the name among other identifiers which can identify both directly and indirectly, depending on the context.[95] In sum, a person can be directly identified not only by name but by reference to another 'unique identifier'.

Consequently, 'indirect' identification refers to the identification through 'unique combinations' of non-unique identifiers. This is what the definition of personal data in part on the modes of identification refers to.[96] An individual can be identified indirectly

> ' . . . by reference to . . . one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.'

A person is indirectly 'identified' when the unique combination of non-unique identifiers is complete and enables to distinguish that person from a group, while when additional information is necessary, that person is indirectly 'identifiable'.

Importantly, whether or not an individual is identified by the available identifiers heavily depends on the context.[97] Similar to how all objective factors need to be considered while assessing whether or not an individual is identifiable,[98] 'the question of whether the individual to whom the information relates is identified or not depends on the circumstances of the case'.[99] For instance, even a name may be insufficient to identify a particular person within a population of a country, when it is a common name, but will likely identify a pupil in a classroom.[100] In the former case, additional information, such as address and date of birth, might be necessary for what will be 'indirect identification'. At the same time, an otherwise non-unique identifier, such as that a person is wearing a black suit, may become unique and hence be sufficient to directly identify a person in a particular context, eg to distinguish one person from the people standing at a traffic light without any additional information.[101]

This results in the meaning of identification as presented in Table 3. A person is 'identified directly', ie distinguished from the group, by name or another unique identifier which is obtained and where no additional information is necessary. A person is directly 'identifiable' when such a unique identifier is not obtained yet, but it is reasonably likely to be obtained. A person is 'identified indirectly' by a unique combination of non-unique identifies which is complete, ie no additional information is needed to identify. A person is 'indirectly identifiable' when such unique combination of identifiers is incomplete and additional information is necessary to be able to distinguish that person.

To illustrate, a website visitor would be 'directly identified' to the website provider by the IP address during the browsing session, because the IP address, whether static or dynamic, is the only and in this case unique identifier that allows the website provider to distinguish one visitor from another. For an example of what would constitute 'directly identifiable', suppose a municipal government, acting within its legal competence, orders all inhabitants of the city of The Hague to stay inside after 20:00. The information that all inhabitants of the city are likely to be inside is information

91 Ibid 13.
92 Ibid 14.
93 Ibid 14.
94 Ibid. This point was also made by eg Borgesius (n 5), but in relation to the meaning of 'identifiable'.
95 I refer to this particular wording of Art 4(1) GDPR here: an individual can be identified 'directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'.
96 WP136 13.
97 '[T]he extent to which certain identifiers are sufficient to achieve identification is something dependent on the context of the particular situation' (ibid).
98 Recital 26 GDPR.
99 WP136 13.
100 Ibid.
101 Ibid 13.

Table 3. Meaning of 'directly or indirectly identified or identifiable'

	DIRECTLY	INDIRECTLY
IDENTIFIED	Distinguished by name or another unique identifier, which *is obtained.*	Distinguished by a unique combination of non-unique identifiers which *is complete* (ie no additional information is necessary)
IDENTIFIABLE	The unique identifier is not yet available, but is *reasonably likely to be obtained.*	By a unique combination of non-unique identifiers where the combination *is incomplete* and additional information is necessary and is *reasonably likely to be obtained.*

that relates to the natural persons who are 'directly identifiable'. While the names of the inhabitants may not be directly available, they are easy to obtain eg from a phone book or a city registry. While this article is reviewed anonymously, a combination of several group characteristics such as current institutional affiliation, gender, nationality, age and field of expertise would be sufficient to specifically pinpoint its author who as a result would be 'identified indirectly'. Suppose one of these characterizations would be missing, making it *prima facie* impossible to single out one specific person in a group, but it was reasonably likely to obtain this additional information. In this case the author would be 'indirectly identifiable'.

Objectively and relatively unique identification

While not discussed in detail by the Working Party, the issue of uniqueness of identification is salient for the meaning of identification under the GDPR. Identification can be either objectively unique, where the chance that another person would have the same identifying attribute(s) is or approaches zero, or it can be relatively unique, where an identifying attribute may not be unique in the world, but is unique in a group or a sample. The threshold the Working Party seems to have adopted is of the relative identification. This follows from the emphasis the Working Party puts on the significance of context for identification. To restate WP136, 'the extent to which certain identifiers are sufficient to achieve identification is something dependent on the context of the particular situation'.[102] A man wearing a black suit by a traffic light is identified by an otherwise not unique attribute (wearing a black suit) in a specific context, ie among a group of passers-by near a traffic light. Another relevant example the WP136 brings, albeit in the context of identifiability, is of key-coded data in research. If codes used to identify research participants are not unique, and the same code (eg 123) is used to distinguish individual participants in different towns and for different years, a possibility of combining the non-unique code with the town and the year will render a participant identifiable[103] and identified if the combination of the code, town and year is complete and held by one actor. That is, according to WP136, an individual will be identified in the sense of the GDPR both by an objectively unique identifier (or a combination of identifiers), and by an identifier that is unique in a particular context, within a sample, or in a group.

Should it be impossible, for technical, logistical, organizational or other reasons, to establish with certainty that information relates to only one unique individual in a given context, the Working Party suggests that the information is still to be regarded as personal data but relating to an 'identifiable' rather than identified natural person. That is, provided the purpose of processing is to identify individuals in a dataset,[104] or the controller cannot establish with absolute certainty that the individuals to whom the data relates cannot be identified.[105]

All five types of identification within the scope of the GDPR

The WP29 approach to identification ensures a far reach of the GDPR as it encompasses the entire integrated typology of identification proposed here, albeit with some reservations. Look-up, recognition, session, and targeting identification do certainly fall within the meaning of identification as distinguishing from the group, as they allow to zoom in on someone as an individual distinct from others, even if the individual remains nameless, or the zooming in is not continuous in time and is limited to duration of a contact or browsing session like in case of targeting and session identification.

102 Ibid.

103 Ibid 19.

104 Ibid 16, 19, see also the video surveillance example.

105 Ibid 17 (the IP addresses example: the controller 'is not in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified').

To illustrate, tracking someone by the IP address constitutes processing of personal data of an identified person, either in the sense of recognition identification if the individual is recognized by his static IP address on a repeated encounter, or in the sense of session identification if the IP address—static or dynamic—is used to distinguish one node in a network from others for the duration of a session. The ability to establish civil identity of the user is irrelevant, and assessing identifiability is not necessary. Using a rich dataset to target tailored content at an individual website visitor is processing of personal data relating to an identified natural person in the sense of targeting identification when the dataset is in use. This is because displaying tailored content different from what others visiting the website see constitutes reaching or distinguishing a person in a group visiting the website at that time. The same dataset when not in use relates to an identifiable person given its purpose to distinguish, ie to identify. Similarly, using rich datasets—stored locally on a user's device or otherwise—to uniquely characterize individuals first in order to put them in larger categories constitutes targeting identification and hence processing of data relating to identified individuals even when these individuals are treated the same, ie as groups, later. This seems to be the case with Google's proposed Federated Learning of Cohorts, or FLoC, alternative to the third-party cookie tracking in interest-based advertising. While the idea is to 'hide individuals "in the crowd"' and 'use on-device processing to keep a person's web history private on the browser',[106] to form cohorts sharing the same interests, similar to the word embedding in the natural language processing, the technique still needs to use the rich browsing history data to map each individual in a multidimensional space to see how they connect to each other.

The status of classification as identification under the GDPR is more complex. Classification on its own is not identification in the sense of the GDPR, even when a broad WP29 approach to identification is adopted. As WP29 has itself pointed out in the context of facial recognition,

> '[a facial] template or set of distinctive features used only in a categorisation system would not, in general, contain sufficient information to identify an individual. It should only contain sufficient information to perform the categorisation (e.g. male or female). In this case it would not be personal data provided the template (or the result) is not associated with an individual's record, profile or the original image (which will still be considered personal data).'[107]

In other words, categorization for as long as it does not uniquely distinguish an individual from a group but simply assigns an individual to a group, does not *prima facie* constitute identification, unless it is applied to an individual identified in other ways, ie through l-, r-, p-, or s-identification. This is similar to the discussion about the status of group profiles as personal data. As among others Koops observes, a group profile becomes personal data when applied to an identified or identifiable person.[108]

However, and importantly, because the WP29 instructs that the possibility to identify heavily depends on a context, classification of a person can become identification in its own right and without the necessary connection to other modes of identification under certain circumstances which make the otherwise non-unique classification a unique identifier. Think of a classification of an individual as a redhead. Although rare, there are still thousands of people with the red hair colour, so classifying someone as a redhead will not be sufficient to distinguish one person from the population of a country, but might be enough in a classroom. Similarly, recall the Working Party example of a person wearing a black suit: an identifier otherwise not unique, but sufficient to distinguish one particular person among the passers-by at the traffic light.

Meaning of identification in CJEU's case law

From *Lindqvist* to *Breyer*: from inconclusive to restrictive interpretation of identification?

Has the CJEU been similarly generous in applying the concept of identification? The Court of Justice ruled on the meaning of personal data in its very first data protection case, *Lindqvist*,[109] and on a number of occasions since then. However, compared to the WP29 opinion, it has not been nearly as articulate on the meaning of the various elements of this concept, including identification. The Court often generally states that the scope of the 1995 Directive—in force when most of the data protection case law was formed—is very wide and the personal data covered by the Directive are varied.[110] The bulk of the relevant cases simply state that a particular type of data is personal, without much explanation or

106 Google's FLoC alternative to the tracking-based targeted advertising in Bindra (n 17).

107 Article 29 Working Party 'Opinion 02/2012 on facial recognition in online and mobile services' (WP192, adopted on 22 March 2012) 4.

108 Bert Jaap Koops, 'Some Reflections on Profiling, Power Shifts, and Protection Paradigms' in Mireille Hildebrandt and Serge Gutwirth (eds), *Profiling the European Citizen: Cross-disciplinary Perspectives* (Springer, Dordrecht 2008) 331; Borgesius (n 5) 260.

109 *Bodil Lindqvist* Case C-101/01 [2003] ECR I-12992, ECLI:EU:C:2003:596.

110 *Österreichischer Rundfunk and Others*, Joined Cases C-465/00, C-138/01 and C-139/01 [2003] ECR I-4989, ECLI:EU:C:2003:294

discussion. Among others, the name of a person but also his telephone coordinates or information about working conditions or hobbies,[111] his address,[112] daily work periods, rest periods and corresponding breaks and intervals,[113] monies paid by certain bodies and the recipients,[114] amounts of earned or unearned incomes and assets of natural persons[115] have been explicitly pronounced to be personal data. Interestingly, *Lindqvist* touches upon the meaning of identification in two paragraphs but is inconclusive, first, on whether or not the data involved is personal because it relates to persons who are identified or identifiable,[116] and whether or not identification in the sense of data protection law can be done via non-name identifiers alone, or in conjunction with a name.[117]

Only relatively recently, did the Court include a more detailed analysis of what particular elements of the concept 'personal data' mean.[118] The case law on the meaning of identification is very limited and inconclusive. In *Scarlet v SABAM*, the Court ruled that the IP addresses of internet users were protected personal data because they 'allow users to be precisely identified',[119] which can be interpreted to mean both that the computer users behind the IP addresses are 'identified' and the IP addresses are the identifiers, and that the computer users are 'identifiable' because the IP addresses make the identification reasonably likely.

In the *Breyer* case the Court focused specifically on the meaning of 'identifiable'. The judgement was welcomed as the assent of the Court to the absolute approach to identifiability in EU data protection law, first declared in Recital 26 of the Data Protection Directive and then adhered to by the WP29.[120] Supporting a broad interpretation of the criterion of identifiability and hence a broad meaning of the concept of personal data, *Breyer* has generally been a positive development in European data protection law. Yet, the role of *Breyer* in establishing the meaning of identification and what it means to be an identified natural person has remained unnoticed.

Davis has recently pointed out that the CJEU in *Breyer* may have invalidated the understanding of identification as distinguishing or being distinguished from a group, advanced by the Working Party. According to Davis, the *Breyer* decision rules out a possibility of direct identification by online identifiers (ie dynamic IP address does not enable the plaintiff to be directly identified). The argument goes: since the Court does not recognize Mr Breyer directly identified by his dynamic IP address, while the very point of IP addresses is to distinguish one website visitor from another, the Court effectively endorses the narrowest understanding of 'identified' as 'identified by name' in the sense of establishing one's civil identity.[121] Indeed, the Court, limiting its considerations on what 'identified' means to one paragraph, concluded that

> it is common ground that a dynamic IP address does not constitute information relating to an 'identified natural person', since such an address does not directly reveal the identity of the natural person who owns the computer from which a website was accessed, or that of another person who might use that computer.[122]

With this conclusion the Court agreed with the referring court[123] and followed the Advocate General:

> [t]he person to which those particulars relate is not an 'identified natural person' [as they]... do not reveal, directly or immediately, the identity of the natural person who owns the device used to access the website or the identity of the user operating the device (who could be any natural person).[124]

This reading of *Breyer* effectively reduces the meaning of identification and 'identified' to the look-up identification, ie by means of identifiers connecting a person to his/her real world identity. This narrow understanding of identification therefore takes out of the protective scope of the GDPR many data-driven practices which have long been assumed to involve personal data processing and thus fall under the GDPR, but which are not

[43]; *Lindqvist*, [88]; and *College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer* Case C-553/07 [2009] ECR I-3889, ECLI:EU:C:2009:293 [59].

111 *Lindqvist* [24].

112 *Rijkeboer* [62].

113 *Worten Equipamentos para o Lar SA v Autoridade para as Condições de Trabalho (ACT)* Case C-342/12 [2013] OJ C225/37, ECLI:EU:C:2013:355 [19], [22].

114 *Österreichischer Rundfunk and Others* [64].

115 *Satakunnan Markkinapörssi and Satamedia* Case C-73/07 [2008] ECR I-09831, ECLI:EU:C:2008:727 [35], [37].

116 The Court makes no distinction between identified and identifiable in its treatment of the case: 'The term ["*identified or identifiable* natural person"] undoubtedly covers the name of a person in conjunction with his telephone coordinates or information about his working conditions or hobbies.' [24].

117 On the one hand, the Court suggests that identification in the sense of the Directive can be achieved also by non-name identifiers: 'act of ... identifying them by name or by other means, for instance by giving their telephone number or information regarding their working conditions and hobbies, constitutes 'the processing of personal data' [27]. On the other hand, the Court directly states that the term 'any information relating to an identified or identifiable natural person' only covers the non-name identifiers in conjunction with the name of a person [24].

118 *YS and others* and *Nowak* focusing on the meaning of "relating to" element of the definition of personal data.

119 *Scarlet v Sabam*, Case C-70/10, ECLI:EU:C:2011:771 [51].

120 Eg Borgesius (n 35).

121 Davis (n 8) 17.

122 *Breyer* [38].

123 *Breyer* [24].

124 AG opinion in *Breyer* [56].

tied to a real-world identity of an individual by his phone number, address, passport number, a name or similar. This includes but is not limited to online behavioural advertising when the data processed does not include real-world identifiers and a person is 'reached' through the online identifiers alone, facial recognition when the facial templates are not associated with the real world identity, individual profiling targeted at a person by means other than offline identifiers, and many others. While in some cases, like in *Breyer*, it may still be possible to argue that the data relates to a person who is 'identifiable' even when the information necessary for the real-world identification is not in the hands of a controller but is 'reasonably likely to be used' for the identification purposes nevertheless, it still does not resolve the resulting gap in legal protection. Indeed, while, as Borgesius correctly argued, the test of identifiability does not require data subjects to be known by name,[125] there must still be a reasonably likely chance of establishing that name or another real-world identifier. In the end of the day, the concept of identifiability is a 'possibility of' identification and its meaning is derived from the meaning of identification. If the latter means real-world identification only, the former must include the possibility of this real-world identification. The possibility to single one out in other ways is not sufficient.

Principle of effective and complete protection and contextual reading of *Breyer*

Yet, the *Breyer* decision has to be read in light of the aim of data protection law to ensure effective and complete protection of data subjects[126] and thus should be construed in such a way that it does not affect the validity of a broader understanding of identification as distinguishing from a group. This reading remedies any restrictive effects on the scope of legal protection.

In order to do so recounting the facts of the case is necessary. The following circumstances gave rise to the case. The websites of the German Federal Government institutions stored the website access logs after the websites have been accessed, which included the name of the web page or file accessed, search terms, the time of access, the quantity of data transferred, whether or not access was successful, and the IP address. Mr Breyer was one of these websites' visitors whose dynamic IP address was retained. He challenged this retention practice in the administrative courts, objecting—on the data protection grounds—to the retention of the IP addresses, unless such retention was necessary to restore the availability of the websites after access failure.[127] The dispute in part concerned whether or not the dynamic IP address constitutes information relating to an identified or identifiable person and thus is or is not personal data. The case went to the court of appeal and finally to the Federal Court of Justice which referred the case to the CJEU. The latter ruled that the IP address does not relate to an identified but to an identifiable natural person.

The key to the alternative interpretation with the effect that an online identifier such as an IP address can identify a person rather than simply render him or her identifiable is in reading the decision with close attention to the context of the case. As the Working Party rightly points out, the context defines whether or not a particular identifier is sufficient to identify a person.[128] In this case, the dispute arose because of the data retention practices of the website owners 'after' the websites were accessed and the browsing session ended. As the Advocate General observes, '[t]he owners of web sites that are accessed using dynamic IP addresses also tend to keep records of which pages are accessed, when and from which dynamic IP address. It is technically possible to retain those records indefinitely after each user terminates his Internet connection'.[129] Mr Breyer objected not to the processing of the dynamic IP addresses per se, eg during the browsing session, but to 'storing, or arranging for third parties to store, *after* consultation of the websites'[130] [emphasis added]. Hence, the question of the referring court also concerned the status of the dynamic IP addresses after consultation of the websites. The referring court submitted that 'the data *stored* does not enable Mr Breyer to be directly identified. . . . [and] [t]he operators of the websites at issue in the main proceedings can identify Mr Breyer only if the information relating to his identity is communicated to them by his internet service provider'.[131] [emphasis added]. The CJEU agreed. However, this does not preclude a conclusion that a website visitor is 'identified' by a dynamic IP address under different circumstances, eg during the browsing session and before the Internet connection is broken. Indeed, when to identify someone means to distinguish that someone from a group, to 'zoom in' on a flesh and blood individual, and implies reaching a person, a website visitor is identified, ie distinguished from other visitors by

125 Borgesius (n 5).

126 *Google Spain* [34], [53]; *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH (Wirtschaftsakademie)* Case C-210/16 [2018] (ECLI:EU:C:2018:388) [28] and *Jehovan todistajat* Case C-25/17 [2018] (ECLI:EU:C:2018:551) [66].

127 *Breyer* [14]–[17].

128 '[T]he extent to which certain identifiers are sufficient to achieve identification is something dependent on the context of the particular situation.' (WP136 13).

129 AG opinion in *Breyer* [4].

130 *Breyer* [17].

131 *Breyer* [24].

means of an IP address and is 'reached' by the website owner in real time when presenting to that visitor the website's content using an IP address. The IP address provides a direct link to a flesh and blood individual who is browsing through the website's content. Under these circumstances, a website visitor is directly identified by the dynamic IP address in the sense of session identification. Once the session is ended and the Internet connection is broken, the retained dynamic IP address is no longer pointing to a specific nod on the network. The direct link with the visitor is severed and additional information is necessary to restore it. This contextual reading of *Breyer* does not effect the validity of the Working Party's understanding of identification as distinguishing a person from the group and preserves a far reach of the GDPR.

This contextual reading of *Breyer* is not only possible as demonstrated above, but also necessary in light of the emerging principle of effective and complete protection of the data subject. The principle was first introduced by the CJEU in *Google Spain*. The principle has been applied since to prevent a narrow interpretation of the concept of a controller and thus against restricting personal scope of the data protection law which would go against the aim to afford a data subject effective and complete protection.[132] As many scholars have argued,[133] the meaning of identification should not be construed narrowly, eg reduced to the look-up identification, because a growing body of invasive data processing practices, such as online advertising, facial recognition, profiling and others, do not have to and often do not rely on the name, address or another real-world identifier. A narrow interpretation of identification and what an 'identified natural person' mean would take those practices and their effects out of the scope of the data protection law and deprive the people affected by them of the GDPR's protection.

Identification as individuation in national case law

The broad interpretation of identification proposed in this article has support in some national case law. Notably, in *Vidal-Hall*, a case concerning processing browser-generated data and cookies by Google, the Court of Appeal of England and Wales ruled that '[i]dentification for the purposes of data protection is about data that "individuates" the individual, in the sense that they are singled out and distinguished from all others'.[134] Thereby the Court recognized recognition identification and rejected a narrow interpretation of identification as by name only: 'It is immaterial that the [browser-generated information] does not name the user. The BGI singles them out and therefore directly identifies them.'[135] For this reason the Court did not find it necessary to consider whether or not the user is 'identifiable', following the Recital 26 test.[136] The same 'individuation' approach to identification was taken by the English court in *Bridges*[137] (although not discussed as relevant on appeal[138]). The case concerned testing of a facial recognition system by the police where CCTV cameras captured facial images of the passers-by within the cameras' range, and first distinguished human faces and then distinguished one face from another, to enable matching the images with biometric templates on the watch lists. The claimant was in the range of the cameras on two occasions and filed a complaint that, among others, his personal data was processed unlawfully, even though he was not matched with the watch list on any occasion. According to the court, there are two routes to argue that personal data is processed. The first route is to be pursued when the data on its own does not qualify as personal data, but additional information can be obtained in future to enable identification. In this case the *Breyer* reasoning is to be followed to establish if identification is reasonably likely and if a natural person to whom the data relates is identifiable.[139] The second route is 'to the effect that a person is sufficiently identified for the purpose of the definition of personal data if the data "individuates" that person'.[140] The second route was followed. The court found that processing of the claimant's image constituted processing of personal data even prior to the matching of the facial images and possible recognition "on the basis that the information recorded by [the facial recognition system] individuates him from all others, i.e. it singles him out and distinguishes him from all others."[141] Since the facial images by themselves directly identified the claimant, the court considered further considerations of the possibility of identification unnecessary.[142] The court effectively

132 *Google Spain* [34], [53] *Wirtschaftsakademie* and *Jehovan todistajat*.

133 Eg Leenes (n 18) and Nissenbaum (n 23).

134 *Vidal-Hall v Google Inc* [2015] EWCA Civ 311 from 114 et seq.

135 Ibid 115.

136 Ibid 124.

137 *R (on the application of Edward Bridges) v The Chief Constable of South Wales Police and Secretary of State for the Home Department* [2019] EWHC 2341 (Admin), 122–25.

138 *R (on the Application of Bridges) v South Wales Police* [2020] EWCA Civ 1058.

139 *R (on the application of Edward Bridges) v The Chief Constable of South Wales Police and Secretary of State for the Home Department* [2019] EWHC 2341 (Admin) at 116–17.

140 Ibid 119.

141 Ibid 122.

142 *Bridges* [123].

recognized targeting identification as a type of identification for the purposes of data protection law.

Conclusion: what this means for data protection

Despite its core role in the EU system of data protection, the notion of identification in the sense of the process and the fact of being identified has been a neglected subject both in data protection law and scholarship. With the primary focus placed on the meaning of identifiability as a legally relevant possibility of identification, it remained unclear the possibility of what exactly is at issue. While Article 29 Working Party interpreted identification broadly, as distinguishing one in a group, this interpretation has been questioned in light of the CJEU decision in *Breyer*, and the uncertainty as to the meaning of identification remained.

This article reduces this uncertainty in two ways. First, it offers an account of what constitutes identification outside of the legal context and proposes an integrated socio-technical typology of identification as a process or result of distinguishing a person in a group. Building on existing socio-technical accounts of identification, the typology distinguishes five identification types: (i) look-up or civil identity identification where persistent identifiers such as a name, passport or social security number, a phone number or address link to a person in a real world, (ii) recognition identification where a person is recognized as known from a previous interaction by a token or another persistent identifier that does not connect to the real-world civil identity, (iii) session identification where a person is reached or linked to an identifier of a limited lifetime for a duration of one interaction, (iv) classification identification where a person is identified as a member of a certain existing group or category by displaying characteristics of that group or category; and (v) targeting identification. The typology has a temporal dimension, in a sense that look-up, recognition and to a limited extent session identification are based on persistent identifiers that enable longitudinal tracking, and classification and targeting identification—if not done together with one of the other types—are transient. The article distinguishes targeting identification as a new identification type, ie selecting in a single moment of time a particular individual from a group as an object of attention or treatment, which can be done either on the basis of a single unique identifier which does not have to be persistent, or on the basis of a rich dataset that uniquely characterizes an individual. Evidence from the literatures on calculated publics, profiling in recommender systems, price and content personalization strongly suggests that targeting as a mode of identification might gain in popularity.

Finally, the article clarifies the legal meaning of identification under the GDPR. If identification—as the Article 29 Working Party would have it—means distinguishing from a group, it unconditionally encompasses the look-up, recognition, session, and targeting types of identification, because they enable reaching, or zooming in on a flesh and blood individual. Whether or not that person is known by name is immaterial. Whether or not classification is identification in the GDPR sense is context-sensitive. Classification only constitutes identification where in a given context, eg a timeframe, a limited space or a group of people, a group characteristic that is otherwise not unique is distinguishing a person as unique in a sample. While the CJEU in *Breyer* seems to have invalidated this approach in favour of name-based identification only, I argue for a contextual interpretation of the decision, which is consistent with the Working Party's position. Such interpretation negates *Breyer's* restrictive potential and does not exclude any of the identification types from the scope of the GDPR.

This has significant implications for data protection law, both short-term on the practical level as well as long-term on the more principal level. Without an ambition of being comprehensive, let me first sketch some illustrative short-term practical consequences. The primary consequence is that the broad interpretation of identification naturally widens the GDPR material scope and leads to a broad application of the GDPR, granting GDPR protections also in the situations where the data subjects are not identified by their civil identities, yet still affected. This includes more conventional cases of identification such as recognition or session identification routinely practiced on the web, but also the more disputed but gaining in popularity cases of the so-called transient data processing where the data subjects are only reached within a brief a moment of interaction with technology, such as the discussed examples of facial detection and analytics in onsite advertising or smart CCTV surveillance.

One can question how significant the role of the concepts 'identification' and 'identified' for the GDPR's material scope would prove to be in practice. Indeed, until now the identifiability test was sufficient for the CJEU to rule that some of the mentioned practices (eg tracking by means of an IP addresses) constitute processing of personal data and thus fall within the scope of data protection law.[143] Similarly, Borgesius argues that targeting practices in the context of behavioural advertising will

143 *Breyer* (n 30) and arguably *Sabam* (n 119).

likely fall within the scope of the GDPR because—although they do not identify a data subject by name—they 'single out' an individual and hence that individual is identifiable.[144] Yet, the impact on the GDPR scope of the interpretation of identification and 'identified' should not be underestimated. While some targeting practices have been included within the scope of data protection law by explicit pronouncement of the CJEU, the status of many other existing and future targeting practices which individuate and have a potential impact on people has not been ruled on yet and is much less certain. It is subject of diverging national interpretations and academic debate.[145] I refer here to the examples of facial detection, emotion recognition and targeted advertising based on transient and edge-, or 'on device', data processing (eg FLoC) discussed in the introduction and throughout this article.[146] Finally, understanding identifiability as the ability to single out might prove to be less than assumed. Scholars seem to be divided as to what identifiability means, and if it includes a possibility of singling out.[147] While Article 29 WP introduced the term 'singling out' in relation to identifiability,[148] the 1995 Directive—in force at the time the relevant opinion was written—did not mention singling out at all. The relation of singling out to the meaning of identifiability under the GDPR is unclear.[149] Given that identifiability draws its meaning from identification, the impact of the later concept on the scope of data protection law and insufficiency to rely on identifiability alone are evident. All these examples and considerations illustrate an urgent need to directly engage with the meaning of identification and 'identified'. This article contains some examples of national courts doing so.[150]

Interpreting targeting as identification in the sense of the GDPR also puts a question mark against some privacy preserving technologies as to what exactly they are preserving. Specifically, the analysis offered in this article counters claims of achieving anonymity that the actors behind some of those technologies make. Consider Google's FLoC alternative to the web-tracking based behavioural advertising. As this article argued, while the objective of the solution is to use on-device data processing to cluster individuals in interest groups rather than target them by tracking their individual behaviour, and thus 'hide individuals "in the crowd"',[151] targeting of the individuals—based on local processing of rich browser data—is a necessary first step of such clustering, and hence the FLoC-based ads are still operating on personal data and do not preserve anonymity in the GDPR sense. Arguing otherwise would not only be against the logic of understanding identification as individuation and distinguishing from a group, but also create an unfair advantage for the big tech companies such as Google who because of their incumbent position within the Internet ecosystem have access to rich datasets that allow them to individuate people first to obfuscate their individualities in clusters later, and imposes an additional burden on smaller less entrenched actors without such access.

Another way how this article is of immediate relevance to data protection practice has to do with the exemptions from an obligation to respect some data protection rights created by Article 11 GDPR, ie exemption from data protection obligations (and rights) if compliance requires identification. According to this provision, in order to encourage data minimization, if the purposes of processing no longer require the data controller to identify a data subject, he shall not be required to maintain the identifying information solely for the purpose of complying with the data protection obligations, including respecting objections to data processing, granting data access, erasure, portability, etc.[152] What it means to identify a data subject is of key importance for the application of this provision. Veale, Binns, and Ausloos bring forward a number of examples where although data subjects can be reached or individuated, eg through WIFI (wireless connection to the Internet) tracking or in case of Apple's voice assistant, the controllers claim that they cannot identify the data subjects and deny access requests.[153] The implication of the present analysis is that any individuation constitutes identification. Hence, Veale, Binns, and Ausloos correctly point out that if a data subject can be reached, even transiently, and despite the fact that the controller may not have the conventional contact data at his disposal to manage data subject requests, the Article 11 exemption does not apply and the rights of access,

144 Zuiderveen Borgesius (n 5) 259 et seq. 'Singling out' is mentioned in the context of the identifiability test in Recital 26 GDPR.

145 Eg as discussed by Davis (n 8) in the context of facial detection.

146 See nn 11–16 and n 106 and the relevant text.

147 Eg George et al argue that data 'only relates to an identifiable person when stored with the purpose of creating profiles that increase an actual risk of re-matching the data' (n 11, 8) but see Borgesius (n 5) arguing that identifiability should be understood as a possibility of singling out.

148 WP136, 12 et seq.

149 In fact, language of Recital 26 GDPR suggests that singling out is one of the 'means reasonably likely to be used . . . to identify the natural person' and consequently that a mere possibility of singling out might not be sufficient to establish identifiability and additional factors need to be taken into account. Davis makes a similar observation (n 8, 13).

150 See nn 136–42 and the relevant text.

151 Chetna Bindra (n 17).

152 Art 11(2) GDPR refers to the rights under arts 15–20.

153 Michael Veale, Reuben Binns and Jef Ausloos, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8(2) International Data Privacy Law 105–23.

objection, portability and other have to be respected. I would add that this is the case for the period that individuation is done for the original purposes of processing. This will most likely require significant efforts in designing the computer infrastructure that facilitates interaction with an individual's device and individuation. For instance, whenever a device is tracked eg using a stable or dynamic MAC address, the device may respond in real time by transmitting objection, access or portability requests, activated in the user settings.[154]

The analysis in this article is also of significance for the status of biometric data as a special category of personal data dependent on the purpose of processing to 'uniquely identify' a data subject,[155] application of the principle of data minimization which restricts processing of data in an identifiable form, of Article 22 GDPR prohibition of automated decision-making when it is based on processing personal data, and for many other matters of the data protection law.

However, next to these practical matters of immediate effect, there is a deeper issue connected to the meaning of identification that is too big to be comprehensively addressed in this article and still needs to be researched. This is the issue of the data protection law's identity: why it exists and what purposes it serves. Identification—along with the other elements of the notion 'personal data'—co-forms the GDPR's material scope and hence reflects the *raison d'être* of data protection. The meaning we attribute to this trigger concept reflects what we see as problematic about data processing: what data protection law is meant to solve, which data or when the data is 'dangerous' and hence when law should intervene. Although all constitute a form of individuation, the five types of identification distinguished in this article are different. As this article pointed out, only look-up identification links to a data subject's real world identity; only look-up and recognition identification and to a limited degree session identification enable tracking of an individual over time and across contexts, while the temporal dimension is irrelevant for the classification and targeting identification. The various types of identification address a variety of practices. This variety may warrant different approaches to legal protection, not all of which are necessarily appropriately accommodated in the GDPR. Any debate on whether or not the data protection's current scope is appropriately drawn, cannot take place before these various problems are disentangled. For instance, if one of the core rationales of the information privacy is to preserve human freedom and autonomy and prevent self-censorship at the fear of being held accountable for one's actions in future, is regulating targeting and classification which do not enable 'remembering' that behaviour needed to achieve this? At the same time, classification and targeting identification, while not relevant for holding data subjects to account for past behaviour, are instrumental for many other instances of automated decision-making and treatment in a broad sense of both terms, including real-time decisions on what product, price or content to display to an individual. This inevitably leads back to the issue of the purpose of data protection law: what is it that we want it to do? Does the desire to include regulation of as much automated decision-making as possible under the umbrella of the GDPR justify stretching the GDPR's scope, and what are the trade-offs? One idea to consider is for regulation of digital harms to start focusing on the potentially problematic practices such as surveillance, profiling and personalization rather than on the proxy concept of personal data, in which case whether or not the affected individuals are identified or identifiable will not be relevant standards to avoid in order to escape regulation. To conclude, the discussion on the meaning of identification in the GDPR is considerably harder than pure legal analysis. It opens many avenues for research on the practical aspects as well as on the principal foundations of data protection law, and requires some normative choices as to what we want data protection law to be.

https://doi.org/10.1093/idpl/ipac013
Advance Access Publication 21 June 2022

154 Ibid.

155 Art 9(1) GDPR.