scripted

Volume 17, Issue 2, August 2020

The Concept of 'Information': An Invisible Problem in the GDPR

Dara Hallinan and Raphaël Gellert***

DOI: 10.2966/scrip.170220.269

Abstract

Information is a central concept in data protection law. Yet, there is no clear definition of the concept in law – in legal text or jurisprudence. Nor has there been extensive scholarly consideration of the concept. This lack of attention belies a concept which is complex, multifaceted and functionally problematic in the GDPR. This paper takes an in-depth look at the concept of information in the GDPR and offers up three theses: (i) the concept of information plays two different roles in the GPDR – as an applicability criterion and as an object of regulation; (ii) the substantive boundaries of the concepts populating these two roles differ; and (iii) these differences are significant for the efficacy of the GDPR as an instrument of law.

Keywords

Data protection; GDPR; information theory; genetic data; artificial intelligence; machine learning

* Senior researcher, IGR, FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH, Karlsruhe, Germany, dara.hallinan@fiz-karlsruhe.de

** Assistant Professor, Faculty of Law, Radboud University, Nijmegen, the Netherlands, r.gellert@jur.ru.nl

1 Introduction

Information is a central concept in data protection law under the General Data Protection Regulation (GDPR).[1] This should be no surprise. Information is, after all the substance, the collection, exchange and manipulation of which, provides the rationale for the existence of data protection law. For a demonstration of the significance of the concept in the GDPR, one needs to look no further than the fact that the concept constitutes a key criterion in the concept of personal data – outlined in art. 4(1) – and therefore plays a defining role in determining whether the law, and all substantive provisions therein, applies at all.

Yet, there is no clear definition of information in European data protection law. There is no definition provided in the text of the GDPR or in prior European Union (EU) data protection law. Nor is there a structured and comprehensive definition provided in relevant jurisprudence. There has been certain scholarly attention paid to the concept in data protection law notably in the excellent work of Bygrave.[2] This work, however, has limitations. The work does not provide a structured approach for the analysis of the functions or boundaries of the concept. Nor does it extensively differentiate between conceptualisations of the concept.

---

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).

2 See: Lee Bygrave, "The Body as Data? Biobank Regulation via the 'Back Door' of Data Protection Law" (2010) 2(1) *Law, Innovation and Technology* 1-25; Lee Bygrave, "Information Concepts in Law: Generic Dreams and Definitional Daylight" (2015) 35(1) *Oxford Journal of Legal Studies* 91-120; Dara Hallinan and Paul De Hert, "Many Have It Wrong – Samples Do Contain Personal Data: The Data Protection Regulation as a Superior Framework to Protect Donor Interests in Biobanking and Genomic Research" in Brent Mittelstadt and Luciano Floridi (eds.), *The Ethics of Biomedical Big Data* (Basel: Springer, 2016), pp. 119-139; Raphaël Gellert, "Data Protection and Notions of Information: A Conceptual Exploration" (2019) SSRN Working Paper, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3284493 (accessed 28 February 2020).

We believe the lack of legal and scholarly attention belies the reality of a concept which is complex, multifaceted and, eventually, functionally problematic in the GDPR. From this perspective, this paper offers an in-depth look at the concept of information in the GDPR and argues three, cumulative, theses:

(1) There are two different roles played by the concept of information in the GPDR: information as an applicability criterion; and information as an object of regulation.
(2) The substantive boundaries of the concepts of information populating these two roles differ – i.e. these are two different concepts, not relating to the same substantive phenomenon.
(3) The substantive differences between these two concepts of information are significant for the efficacy of the GDPR as an instrument of information law.[3]

The paper begins by sketching the two roles played by the concept of information in the GDPR (section 2). The paper then advances a conceptual framework – built on three axes – for identifying the substantive boundaries of conceptualisations of information in each of these two roles (section 3). Using this framework, the paper then maps the substantive boundaries of the two concepts: first, the concept of information as an applicability criterion (sections 4-8); second, the concept of information as an object of regulation (sections 9-12). Building on this

[3] The aim of this paper is to sketch the contours of an important, albeit largely ignored, topic of research in data protection law under the GDPR: the concept of information. In this regard, the paper should not be taken as offering a final authoritative position on the exact functions, boundaries or significance of the concept in the GDPR, nor as offering specific accounts of the extent of problems caused by the concept for the function of the GDPR, nor as suggesting that the problems caused by the concept are more important than, or replace, problems caused by other concepts in the GDPR. Further clarification of such complicated definitional and relational issues requires, and deserves, much further research.

mapping, the paper highlights the substantive differences between the two concepts (section 13). The paper then shows how divergences between the two concepts are problematic for the GDPR as an instrument of information law (sections 14-16). Finally, the paper considers the legal options available for addressing these problems (section 17).

We begin by sketching our first thesis: *There are two different roles played by the concept of information in the GPDR.*

2 Two Different Roles for the Concept of Information in the GDPR

Considerations of the role of the concept of information in data protection law – including in the GDPR – have tended to explicitly identify only one role: information as an applicability criterion (role 1).[4] This is understandable, as the concept explicitly appears in legal text only in this role. We, however, would suggest that the concept also plays a second role: information as an object of regulation (role 2). Below, we sketch the function of each of these roles in the GDPR.

In the first role, as an applicability criterion, the concept of information functions to define whether the GDPR can apply *rationae materiae*. Art. 2 of the GDPR outlines the law's scope. Art. 2(1) elaborates a key applicability criterion – the concept of personal data: "This Regulation applies to the processing of personal data wholly or partly by automated means." Art. 4(1) provides a definition for personal data in which information is listed as an explicit substantive criterion for the existence of personal data: "'personal data' means any *information* relating to an identified or identifiable natural person ('data

[4] See, for example: Bygrave, "The Body as Data?", *supra* n. 2. Although Bygrave does indicate the existence of a second role, this is not made explicit as an object of definition and analysis.

subject')" (emphasis added). Thus, the presence or absence of information determines which substances can, or cannot, be personal data and to which the GDPR and its substantive provisions can apply.[5]

In its second role, as an object of regulation, the concept of information functions as a substance around which the substantive principles of the GDPR were designed, and in relation to which these principles will act – much as, for example, medical devices are the object of the EU medical devices law.[6] This concept is implicit in the GDPR. Specifically, this concept must have taken some form in the mind of the legislator for the legislator to have engaged in the choice and design of substantive provisions. For example, in art. 15 of the GDPR – concerning data subjects' rights of access their personal data – the data subject has the right to obtain, from the data controller, a copy of their personal data. To provide this copy, the controller must perform a set of actions regarding the substance of information. That such a provision appears in the GDPR means the legislator must have had some image of the characteristics of the substance of information, and as to how a controller might engage with it.

Superficially, it would make sense that the concepts of information occupying these two roles would converge on the same substantive phenomenon i.e. the concepts would have the same substantive boundaries. A closer look, however, reveals reason to think otherwise. As will be discussed in the next section, there are numerous concepts of information and the boundaries of these concepts can differ significantly. These differences often result from the different functions the concept of information plays in the context in which it is employed.

---

5 Article 29 Data Protection Working Party, "Opinion 4/2007 on the Concept of Personal Data" (WP136, 2007), pp. 6-9, available at https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf (accessed 3 February 2020).

6 Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC.

In this regard, there is a clear difference between the function of the concept of information in each of its two roles in the GDPR. As an applicability criterion, the concept performs a normative function defining whether a substance qualifies for protection at all.[7] As an object of regulation, the concept plays a descriptive function describing a substance with a specific set of properties around which to legislate, and subsequently, act.

The previous section sketched our first thesis that the concept of information plays two roles in the GDPR. We highlighted the following two roles:

(1) Information as an applicability criterion (role 1).
(2) Information as an object of regulation (role 2).

This section also suggested there is reason to think the substantive boundaries of the concepts occupying the two roles may not converge on the same substantive phenomenon. Against this background, we thus move to elaborate our second thesis: *The substantive boundaries of the concepts of information populating these two roles differ.* The first step in demonstrating this thesis is to elaborate a general framework for the structured mapping and differentiation of concepts of information.

3 A Framework for Mapping the Two Concepts of Information in the GDPR

A comprehensive mapping of the boundaries of a legal concept ideally follows within a structured conceptual framework outlining the range of possible dimensions of the concept. Identifying such a framework would normally follow

[7] As Taylor puts it, the concept of personal data functions as a "gateway to the application of data protection principles." Mark Taylor, *Genetic Data and the Law: A Critical Perspective on Privacy Protection* (Cambridge: CUP, 2012), p. 77.

a consideration of relevant law and jurisprudence. In relation to the concept of information in the GDPR, however, there are insufficient legal resources to identify such a framework.[8] To overcome this obstacle, we construct a structured conceptual framework via a consideration of the phenomenology of information, not from a jurisprudential, but from a general perspective. At the highest level of abstraction, information is a resource for the resolution of uncertainty. In this regard, myriad disciplines have adopted concepts of information. In doing so, however, each discipline – depending on the purpose of the concept in the discipline – has defined the concept differently. We identify a set of three key axes differentiating concepts of information across disciplines.[9] Taken together, these axes provide a structured conceptual framework within which to map concepts of information in the GDPR.

Axis 1: the degree to which information must be semantic relate to meaning in the world. Not all conceptualisations of information require information to convey meaning about the world. Mathematical concepts of information, for example, focus on the probabilistic relationships between systems regardless of semantic content. The typical example is Shannon

[8] The most systematic elaboration of the concept of information in EU data protection law is offered in the Article 29 Working party, "Opinion on the concept of personal data", *supra* n. 5, pp. 6-9. There are issues with the approach in this Opinion, however. Three stand out. First, the Opinion only considers information in relation to the concept of personal data and thus in its role as an applicability criterion – the limitation of this scope will become clear as this paper progresses. Second, how the Article 29 Working Party drew up their schema for analysing the concept is unclear and eventually, misses certain key aspects of the concept such as the relationship between information and human cognition. Finally, the Opinion contains contradictions and lacks clarity.

[9] We view these axes, in no way, as being exhaustive or definitive. We appreciate the possibility for the addition of further significant axes, as well as the possibility for alternative approaches to the conceptualisation of axes. We hope that other scholars may do just this. The selection of axes relies, in particular, on the helpful breakdowns of concepts of information by Zins. Chaim Zins, "Conceptual approaches for defining data, information and knowledge" (2007) 58(4) *Journal of the Association for Information Science and Technology* 479-493, pp. 487-489. The selection of axes also relies on prior knowledge of the background of EU data protection law, its *modus operandi* and consequent assumptions as to the types of axes which might be likely to provide fruitful points of reference for analysis.

information, which concerns the statistical properties of systems and the correlation between the states of two systems – regardless of semantic content of states.[10] In turn, not all semantic information corresponds to meaning in the same way. Most importantly, information may differ in the degree of structuring required to convey meaning to an agent. Information may be deliberately structured to convey meaning frictionlessly – for example a factual sentence – or information may be less unstructured, requiring the addition of more, or less, complex interpretative frameworks to extract meaning.[11]

Axis 2: the degree to which information must be stored and transferred within, and across, specific media. Certain conceptualisations of information focus on the requirement for information storage and transfer within specific media. Certain definitions in computer science, for example, may insist on the necessity for information storage and transfer in computer media or at least in human-created media.[12] There are other definitions of information, however, which cast the net wider. Certain philosophical definitions point to the feasibility of naturally occurring information.[13] This information is stored in naturally occurring physical phenomena. The typical example are the rings located in tree trunks. These rings exist independently of human storage media – and even of

[10] See: Claude Shannon and William Weaver, *The Mathematical Theory of Communication* (Chicago: University of Illinois Press, 1949). Several other such approaches are also identifiable.

[11] The reader might, at this point, wonder why we have not simply used the term data when talking about unstructured information. This is a linguistic choice to avoid confusion later in the paper. Specifically, the terms information and data are used almost interchangeably in EU data protection law. Eventually, EU data protection law is unconcerned with data as such – at least in the sense the term is used in other contexts – but rather only with the potential information which may be contained within data. In order to avoid terminological confusion, we thus only talk about different degrees of structure in information.

[12] *Supra* n. 9, p. 488.

[13] Neil Manson, "The Medium and the Message: Tissue Samples, Genetic Information and Data Protection Legislation" in Heather Widdows and Caroline Mullen (eds.), *The Governance of Genetic Information: Who Decides?* (Cambridge: CUP, 2009) pp. 15-36; 20.

human observation. Yet, the rings correlate with the age of the tree and therefore may be considered in terms of information.[14]

Axis 3: the degree to which information must relate to human cognition. Certain conceptualisations of information focus on some degree of human cognition in the creation or perception of information. These definitions tend to correspond, in terms of use, with those requiring information to be stored or transferred on specific media. For example, the International Organization for Standardization – in defining information technology vocabulary – suggests information to be: "knowledge which reduces or removes uncertainty."[15] Knowledge requires cognition. Other definitions are human cognition ambivalent. For example, biological conceptualisations of information regard the function of DNA – both in terms of inheritance and translation between genotype and phenotype – in terms of information.[16] DNA information is created, and operates independently of human cognition.

We now move to map the concept of information in each of its two roles in the GDPR. The mapping process for the concept in each role involves two steps:

(1) Provide an overview of the background to the concept of information to offer perspective and orientation to the mapping process.
(2) Map the concept of information against each of the three axes of differentiation outlined in this section, above.

Both steps are applied first to the concept of information as an applicability

---

[14] Luciano Floridi, *The Philosophy of Information* (Oxford: OUP, 2011), p. 43.

[15] International Organization for Standardization, *ISO/IEC 2382:2015 Information technology — Vocabulary* (ISO/IEC 2382:2015, 2015), available at https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:ed-1:v1:en (accessed 3 March 2020).

[16] See, for example, Paul Griffiths and Karola Stotz, *Genetics and Philosophy: An Introduction* (Cambridge: CUP, 2013), pp. 153-158.

criterion (role 1) and then to the concept of information as an object of regulation (role 2).

4 Providing an Overview of the Background of Information as an Applicability Criterion (Role 1)

The concept of information as an applicability criterion has a long and stable history in European data protection law. This history stretches back to the earliest international instruments of data protection law with European relevance. The concept was evident as an applicability criterion in the Organisation for Economic Co-operation and Development Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980) as well as in the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981) – the concept also appears in the updated versions of these instruments.[17] The concept was then retained, in fundamentally unaltered form, in both Directive 95/46 – the Data Protection Directive and the forerunner to the GDPR – and in the GDPR.[18]

As discussed above – in section 2 – arts. 2(1) and 4(1) recognise the concept of information as one criterion, amongst a set of applicability criteria, all of which must be fulfilled for the GDPR to apply.[19] The criterion of information, however, is conceptually distinct from other art. 2(1) and 4(1) criteria. The criterion applies

[17] Organisation for Economic Co-operation and Development Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [1980], art. 1(b); Organisation for Economic Co-operation and Development Guidelines on the Protection of Privacy and Transborder Flows of Personal Data [2013] art. 1(b); Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [1981], art. 2(a); Council of Europe Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [2018].

[18] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, art 2(a). Even though some earlier national statutes relied upon the much narrower definition of "biographical information" – see section 9.

[19] *Supra* n. 5, pp. 6-9.

to a basic class of substances to which the GDPR can apply regardless of context – i.e. a substance either is, or is not, information, regardless of subsequent elements of context. All other art. 2(1) and 4(1) criteria are then context-dependent. The applicability of other art. 4(1) criteria defining the concept of personal data – "relating to an identified or identifiable natural person" – are contingent on the presence of a contextually defined link between information and a specific individual. The applicability of the art. 2(1) criteria of "processing…wholly or partly by automated means" are contingent on a set of actions being done to information.

Given this conceptual specificity, the substantive content of the concept can be considered independently from other art. 2(1) and 4(1) applicability criteria. This possibility has been explicitly recognised in jurisprudence. The Article 29 Working Party, for example, in their Opinion on the Concept of Personal Data, devote a specific section to the consideration of the concept of information apart from other art. 4(1) criteria.[20] Equally, the CJEU, when considering the applicability of the concept of personal data in the *Nowak* case, considered the concept of information as an applicability criterion separately from other art. 2(1) applicability criteria.[21] Indeed, recognising the independence of other art. 2(1) and 4(1) applicability criteria also has a long scholarly tradition. Consider, for example, the independent scholarly analyses of the art. 2 concepts of "related to" and "identifiability."[22]

The concept has always been intended to be understood and interpreted considering its function in EU data protection law. The base rationale of EU data

---

[20] *Supra* n. 5, p. 5.

[21] *Peter Nowak v Data Protection Commissioner*, C-434/16, [2017] ECLI:EU:C:2017:994, paras 33-35 (hereinafter *Nowak*).

[22] See, for example, Worku Gedefa Urgessa, "The Protective Capacity of the Criterion of 'Identifiability' under EU Data Protection Law" (2016) 2(4) *European Data Protection Law Review* 521-531, p. 521.

protection law, as outlined in art. 1(1) of Directive 95/46 – substantively unchanged in the GDPR – was to: "protect the fundamental rights and freedoms of natural persons." The aim of data protection, broadly put, is thus to provide protection whenever individual rights are threatened in the information society. Accordingly, since its first use in EU data protection law, the concept was intended to be interpreted broadly and flexibly to ensure data protection law applied whenever its base rationale was fulfilled. In this regard, in the *travaux preparatoires* for Directive 95/46 the Commission of the European Communities explicitly recognised the need for a broad, flexible definition of personal data – and therefore of information as one of its constituent criteria: " 'Personal data'. As in Convention 108, a broad definition is adopted in order to cover all information which may be linked to an Individual."[23]

The need for a flexible and broad approach to the interpretation of the concept of information as an applicability criterion has been reaffirmed in subsequent jurisprudence. The Court of Justice of the European Union (CJEU), for example, in the recent case of *Nowak*, held: "The use of the expression 'any information' in the definition of the concept of 'personal data', within art. 2(a) of Directive 95/46, reflects the aim of the EU legislature to assign a wide scope to that concept…potentially encompass[ing] all kinds of information…provided that it 'relates' to the data subject."[24] In turn, the Article 29 Working Party, in their Opinion on the Concept of Personal Data, observed: "The term 'any information' contained in the Directive clearly signals the willingness of the legislator to

[23] Commission of the European Communities, *Commission Communication on the protection of individuals in relation to the processing of personal data in the Community and information security* (COM(90) 314 final, 1990), p. 19, available at http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:51990DC0314&from=EN (accessed 4 March 2020).

[24] *Nowak* n. 21, para. 34.

design a broad concept of personal data. This wording calls for a wide interpretation."[25]

Against this background, we now move to map the substantial boundaries of the concept of information as an applicability criterion against each of the three differentiating axes of the structured conceptual framework – outlined in section 3. We consider the substantive boundaries of the concept, along each axis, from two perspectives:

(1) In terms of teleology – in light of the function of the concept in relation to the basic rationale of the GDPR.
(2) In terms of whether there are further refinements of the concept identifiable in jurisprudence.

5 Mapping the Concept of Information as an Applicability Criterion in Terms of the Relationship between Information and Meaning (Role 1, Axis 1)

From a teleological perspective, the concept of information as an applicability criterion can relate only to semantic information. The purpose of data protection law under the GDPR is to protect individuals in relation to concerns around the use of their information in social contexts – by bureaucracies and by economic actors. Such social concerns arise only regarding semantic information. Social concerns arise only concerning power relations created by other actors knowing – or potentially knowing – something about an individual with social relevance. In this regard, non-semantic concepts of information are – if not *prima facie* excluded – largely meaningless. As Bygrave generally observes: "Information usually denotes a form of semantic content in law…. Law is primarily concerned

[25] *Supra* n. 5, p. 6.

with regulating human relations; therein, the production and exchange of meaning play a key role."[26]

In this regard, from a teleological perspective, the concept should encompass all semantic information regardless of the degree to which interpretation is still required to produce meaning to an agent. The degree of structuring of information in terms of meaning is not definitive of the existence, or degree, of risks to individuals' rights and freedoms pertaining to information processing. Accordingly, the concept covers unstructured information which requires further interpretation to produce meaning to an agent, all the way up to clearly structured facts. This was demonstrated in the European Court of Human Rights' (ECtHR) *Marper* case. In this case, the Court recognised the risks to rights and freedoms relating to the processing of unstructured genomic information – the raw genomic code. The Court recognised such processing as: "interfering with the right to respect for the private lives of the individuals concerned."[27]

---

[26] Bygrave, "Information Concepts in Law" *supra* n. 2, p. 112. See also: Raphaël Gellert, "Organising the regulation of algorithms: comparative legal lessons" (2019) Presentation given at the TILTing 2019 Conference.

[27] *S. and Marper v United Kingdom*, app. no. 30562/04 and 30566/04, [2008], para. 73 (hereinafter *Marper*). We recognise that the case did not explicitly use the term "unstructured genomic information." The case did, however, deal with cellular samples which the Court recognised as being of significance in relation to the individual's private life as these contain the raw genomic code – in DNA. This is raw form genomic information which requires further analysis through an interpretative framework – provided by genetic science – to extract information with social significance about an individual. Hence, the raw genomic code might be referred to as unstructured genomic information. For example, in order to extract information from an individual's genome as to whether that individual has a genetic predisposition to contract Huntington's disease, an interpretative framework based around the detection of a mutation in the HTT gene would need to be applied. See: David Craufurd et al., "Diagnostic genetic testing for Huntington's disease" (2015) 15(1) *Practical Neurology* 80-84, p. 80. Indeed, the Court specifically recognised the significance of the raw genomic code for individuals' private life due to the possibility to subject the code to different types of interpretative framework to produce different types of socially significant factual information about those individuals – and indeed their relatives. The Court stated, for example: "In addition to the highly personal nature of cellular samples, the Court notes that they contain much sensitive information about an individual, including information about his or her health. Moreover, samples contain a unique genetic code of great relevance to both the individual and his relatives. In this respect the Court concurs with the opinion expressed by

Jurisprudence does little to further delimit the forms of semantic information covered by the concept of information as an applicability criterion. In fact, jurisprudence has only hinted at limitations on the range of semantic information covered by the concept in one case. This case concerned whether opinions and inferences – extrapolations about individuals from other available information – qualify as the subject of data protection law. This doubt emerged after the 2014 CJEU *Y.S. and M. and S* cases. In the case, the Advocate General – in an opinion followed by the Court – concluded: "only information relating to facts about an individual can be personal data."[28] The suggestion that the concept only relates to facts, however, was expunged in the subsequent 2017 CJEU *Nowak* case. In this case, the CJEU explicitly clarified: "all kinds of information…also subjective, in the form of opinions and assessments [are personal data]."[29]

6 Mapping the Concept of Information as an Applicability Criterion in Terms of the Relationship between Information and Media (Role 1, Axis 2)

From a teleological perspective, the concept of information as an applicability

Baroness Hale in the House of Lords (see paragraph 25 above)" – the opinion with which the Court was agreeing was the following: "the retention of both fingerprint and DNA data constituted an interference by the State in a person's right to respect for his private life and thus required justification under the Convention. In her opinion, this was an aspect of what had been called informational privacy and there could be little, if anything, more private to the individual than the knowledge of his genetic make-up." Paras 71 and 25.

[28] *Opinion of Advocate General Sharpston in YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S,* Joined Cases C-141/12 and C-372/12, [2013], para. 56. See also Sandra Wachter and Brent Mittelstadt, "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Inferences and Big Data" (2019) 2019(2) *Columbia Business Law Review* 494-620, pp. 521-531.

[29] *Nowak* n. 21, para. 53. *Nowak* is welcome for consistency and doctrinal integrity. In terms of consistency, it would be hard to reconcile the possibility for uninterpreted datasets, as well as facts, to fall within the scope of information as an applicability criterion whilst opinions and inferences could not. In terms of doctrinal integrity, the goal of EU data protection law is doubtless relevant in relation to opinions and inferences. See also Dara Hallinan and Frederik Zuiderveen Borgesius, "Opinions can be incorrect (in our opinion)! On data protection law's accuracy principle" (2020) *International Data Privacy Law* (forthcoming).

criterion is ambivalent as to the medium of information storage or transfer. In terms of purpose, the concept aims to encompass all semantic information relating to an individual, the processing of which might constitute a risk for that individual. As the teleology of the concept relates to the semantic content of information, the media of storage and transfer are incidental. Accordingly, from this perspective, there is no limitation on the media which may be encompassed by the concept. The concept can encompass information stored and processed in computer-based information processing systems, information stored and processed in other artificial man-made systems and even information stored in naturally occurring media – for example DNA stored in a biological sample.[30]

The ambivalence of the concept to the media of storage and transfer is generally affirmed in jurisprudence. There is only limited CJEU case law on the matter. Yet, where the issue has been discussed, the Court has always recognised the concept extends to cover the information storage and transfer media in question. In the recent CJEU cases of *Ryneš* and *Buivids*, for example, the Court highlighted the concept encompasses information stored and processed in sound and image form.[31] More extensive consideration comes from the Article 29 Working Party, in their Opinion on the Concept of Personal Data. In the Opinion,

---

[30] See Dara Hallinan, *Feeding Biobanks with Genetic Data: What role can the General Data Protection Regulation play in the protection of genetic privacy in research biobanking in the European Union?* (Brussels: Vrije Universiteit Brussel, 2018), p. 99.

[31] In the case of *Ryneš,* the Court stated: "It should be noted that, under Article 3(1) of Directive 95/46, the directive is to apply to 'the processing of personal data wholly or partly by automatic means, and to the processing otherwise than by automatic means of personal data which form part of a filing system or are intended to form part of a filing system'…Accordingly, the image of a person recorded by a camera constitutes personal data within the meaning of art. 2(a) of Directive 95/46 inasmuch as it makes it possible to identify the person concerned." *František v Úřad pro ochranu osobních údajů,para,* Case C-212/13, [2014] ECLI:EU:C:2014:2428, paras 20-22.
In the case of *Buivids,* the Court stated: "In the present case, it is apparent from the order for reference that it is possible to see and hear the police officers in the video in question, with the result that it must be held that those recorded images of persons constitute personal data within the meaning of art. 2(a) of Directive 95/46." *Sergejs Buivids,* Case C–345/17, [2019] ECLI:EU:C:2019:122, para. 25.

the Working Party made the following statement: "Considering the format or the medium on which…information is contained, the concept of personal data includes information available in whatever form, be it alphabetical, numerical, graphical, photographical or acoustic, for example. It includes information kept on paper, as well as information stored in a computer memory by means of binary code, or on a videotape, for instance."[32] Despite general affirmation in jurisprudence, however, doubt has been raised in one specific case.

This case concerns whether information stored in biological form, in DNA in human biological samples, can fall within the concept of information as an applicability criterion under the GDPR. Doubt emerges on the back of the same Article 29 Working Party Opinion discussed in the preceding paragraph. In this regard, the Working Party stated: "Human tissue samples (like a blood sample) are themselves sources out of which [information is] extracted, but they are not [information] themselves."[33] This statement seems conclusive. Yet, the Article 29 Working Party position is not supported by clear substantive argumentation. In fact, a deeper investigation reveals strong evidence that information stored in biological form, in DNA in human biological samples, should be regarded as falling under the concept of information as an applicability criterion in the GDPR. As this issue is seldom-discussed, the next section will outline the argumentation supporting the position in more detail.

7 Mapping the Concept of Information as an Applicability Criterion in Terms of the Relationship between Information and Media: The case of information in DNA

The argumentation rests on three pillars: first, the teleological legitimacy of the

[32] *Supra* n. 5, p. 7.
[33] *Supra* n. 5, p. 9.

position – touched on above, now elaborated in detail; second, the legal-technical legitimacy of the position; and third, the jurisprudential support for the position.

In the first instance, there is a strong case for the teleological legitimacy of the position. There are many situations in which biological samples are collected, stored and processed for the genomic data they contain – for example biobanking. In these cases, storage and transfer of information in biological form – in DNA – is practically equivalent to the storage and transfer of sequenced genomic data. As a result, the processing of biological samples engages an equivalent set of rights to the processing of sequenced genomic data. As Bygrave observes, in such contexts: "it is increasingly difficult, in practice, to distinguish between data/information and their biological carriers…there is frequently an intimate link between biological samples and the information they generate."[34] Thus, if the concept of information as an applicability criterion should be interpreted broadly to apply to all types of information whenever individuals' rights in information are at risk, and – as discussed in section 5 in relation to the *Marper* case – such risks are engaged by the processing of individuals' genomic information, then the concept should surely also apply to information stored in DNA in biological samples.

In turn, there are no clear legal-technical obstructions which can be raised against the position. Two forms of legal-technical argument against the position have been put forward. These, however, are both flawed. First, an argument has been put forward that biological samples – and therefore the DNA contained therein, cannot technically qualify as information at all. As Nys put it: "data are representations of reality, whereas human biological materials are real themselves."[35] Contrary to Nys' assertion, however, the dominant

[34] Bygrave, "The Body as Data?", *supra* n. 2, p. 20.

[35] Herman Nys, "Report on the Implementation of Directive 95/46/EC in Belgian Law" in Deryck Beyleveld, David Townend, Ségolène Rouillé-Mirza and Jessica Wright (eds.), *Implementation*

characterisation of DNA in biological samples is as information. DNA is conceptualised as information in popular understanding, as well as in the genetic sciences.[36] Indeed, such is the proximity of the comparison, DNA has even been put forward as an alternative to computer-based information storage.[37] If DNA cannot be information, should an extensive file on an individual stored in the medium of DNA not qualify as information either?

Second, an argument has been put forward that the concept of information as an applicability criterion in EU data protection law may have been built around a concept of information drawn from informatics, and that such a discipline-specific concept cannot support the inclusion of biological samples.[38] There is, however, no evidence that such a discipline-specific concept of information was intended as a template for the concept in the GDPR, or in any prior instrument of EU data protection law – in either the legal texts themselves or in the *travaux préparatoires*.[39] Even if such a discipline specific concept had been used as a template, DNA in biological form could still be conceived of as information. In Zins' work on the concept of information in informatics, for example, several definitions of the concept of information in informatics are

of the Data Protection Directive in Relation to Medical Research in Europe (Aldershot: Ashgate, 2004) pp. 29-41, p. 41.

36 See the discussion as to how biological samples, and the DNA they contain are conceptualised in popular metaphors and in genetic science in Hallinan and De Hert, "Many Have It Wrong", *supra* n. 2, pp. 131-133.

37 See, for example, George Church, Yuan Gao, and Sriram Kosuri, "Next Generation Digital Information Storage in DNA" (2012) 337(6102) *Science* 1628, p. 1682.

38 See, for example, the recognition of these arguments in Bygrave, "The Body as Data?", *supra* n. 2, pp. 14-16.

39 For a more extensive discussion as to the lack of proof of any intention to use a concept of information in informatics as the template for the concept of information in the GDPR, see Hallinan and De Hert, "Many Have It Wrong", *supra* n. 2, pp. 133-134.

identifiable which encompass DNA.[40] The most authoritative definition of information in informatics, offered in ISO 2382-1, can also encompass DNA.[41]

Finally, the position has growing jurisprudential support. In this regard, we would highlight the existence of three decisions before the ECtHR in which DNA was explicitly recognised in terms of personal data – and therefore in terms of information. The most well-known of these cases is the *Marper* case – already discussed above. In this case, the Court explicitly recognised that: "cellular samples [as carriers of DNA], constitute personal data."[42] There are, however, two other, more recent, cases, in which the Court has reiterated this position. In both *Gaughran* and *Trajkovski and Chipovski*, the Court stated: "The Court notes that…DNA material is personal data."[43] In principle, jurisprudence from the ECtHR – as a Court capable of making binding decisions in relation to Member States – should be regarded as having greater legal significance than competing claims in an Article 29 Working Party Opinion.[44]

---

40 *Supra* n. 9, pp. 485-486.

41 The International Organization for Standardization define data as: "A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing…Data can be processed by humans or by automatic means." *Supra* n. 15. For an extensive discussion of the way in which DNA falls within the definition of information in ISO 2382-1, see Hallinan and De Hert, "Many Have It Wrong", *supra* n. 2, p. 134.

42 *Marper*, *supra* n. 27, para. 68.

43 *Gaughran v United Kingdom*, app. no. 45245/15, [2020], para. 63; *Trajkovski and Chipovski v North Macedonia*, app. no. 53205/13 and 63320/13, [2020], para. 43.

44 It should be noted that, in the case, the Court did highlight that cellular samples may not always constitute personal data. The Court recognised that biological samples would only constitute personal data if they were able to fulfil all the criteria of the concept of personal data outlined in the Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981): "The Court notes at the outset that all three categories of the personal information retained by the authorities in the present case…DNA profiles and cellular samples, constitute personal data within the meaning of the Data Protection Convention as they relate to identified or identifiable individuals. The Government accepted that all three categories are "personal data" within the meaning of the Data Protection Act 1998 in the hands of those who are able to identify the individual." *Marper* n. 27, para. 68. With these observations, the Court suggested that biological samples need not always be identifiable and that, accordingly, they need not always be personal data. We would highlight, however, that this recognition does not alter the fact that, by suggesting cellular

It may be argued that these assertions may not reflect the Court's considered position on the definition of information as a constituent criterion of personal data and should thus be taken with caution. This argument pivots on the fact that the Court's statements in each case were brief and not supported by substantive argumentation.[45] There are two reasons, however, that this argument cannot be accepted. First, the statements in the latter two cases are key to the Court's subsequent argumentation. In both cases, the statements provide the base justification for the finding of an interference with the applicants' right to private life. It seems highly unlikely the Court would build its legal reasoning around an unconsidered position. Second, the statements concern a concept – personal data – with an extensive history in Council of Europe law and ECtHR jurisprudence.[46] Recall the concept already appeared as an applicability criterion in the Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981) – a concept on which the concept of information as an applicability criterion in the GDPR is based.[47] It seems highly unlikely the Court failed to recognise the history or significance of

samples can, in some cases, be personal data, the Court recognised that cellular samples will always – provided they contain DNA – constitute information. As discussed above, in section 4, if a substance can, in principle, fulfil the information criterion of personal data, but then cannot fulfil the other criteria of personal data – for example if a substance is not identifiable in a specific case – this does not have the effect of altering its classification as information. This only has the effect of altering the substance's context specific classification as personal data by virtue of its failure to fulfil other criteria.

45 In relation to the statements in the *Marper* case, for example, Bygrave sums up the sentiments behind this argument as follows: "The finding by the ECtHR that samples constitute personal data is…remarkable for its brevity…in formulation and…reasoning." Bygrave, "The Body as Data?", *supra* n. 2, p. 8.

46 See, for example, in this overview the extensive history of cases concerning personal data before the ECtHR: European Court of Human Rights, *Personal data protection* (2020), available at https://www.echr.coe.int/Documents/FS_Data_ENG.pdf (accessed 24 February 2020).

47 Art. 2(a) of the Convention reads: "'personal data' means any information relating to an identified or identifiable individual ('data subject')." Compare this with the definition provided in art. 4(1) of the GDPR: "'personal data' means any information relating to an identified or identifiable natural person ('data subject')."

the concept, or acted carelessly in relation to the concept, when making its comments.

8 Mapping the Concept of Information as an Applicability Criterion in Terms of the Relationship between Information and Human Cognition (Role 1, Axis 3)

From a teleological perspective, the concept of information as an applicability criterion engages human cognition in an indirect way. Human cognition need not play a direct role in the creation or perception of information for information to attain social significance – and thereby pose a risk in terms of individuals' rights. For example, the sequencing and digital storage and transfer of an individual's genome could happen automatically, without any human scrutiny or interrogation of the pertinent information. Yet, none would argue the processing of a sequenced genome poses no risk to individuals' rights.[48] From a teleological perspective, however, human cognition does need to play an indirect role in setting the information processing context – programming computer systems, for example – and in setting the processing agenda. It is impossible to imagine a situation in which semantic information could obtain social significance, and thus pose a risk to rights, without human cognition playing some role in setting the processing context.

The distinctions outlined in the paragraph above are supported in jurisprudence. There has been little jurisprudence specifically dealing with the relationship between human cognition and the concept of information as an applicability criterion. The matter has, however, received certain indirect consideration. In this consideration, jurisprudence has provided two significant clarifications. First, jurisprudence has generally clarified that the concept is

[48] See for example *Marper*, *supra* n. 27, as discussed in sections 6 and 7.

ambivalent as to whether human cognition has played an active role in the creation or perception of information. In the CJEU *Digital Rights Ireland* case, for example, the Court recognised that systems which automatically store and retain information – independent of human cognition – constitute systems which process personal data. The Court thereby confirms that information processed in such systems can fall within the concept of information as an applicability criterion.[49]

Second, jurisprudence has clarified that the concept of information as an applicability criterion is not only ambivalent as to whether human cognition has played a role in the creation and perception of information processed, but is also ambivalent as to whether human cognition has played a role in determining the semantic content of information processed. In their Opinion on Online Behavioural Advertising, for example, the Article 29 Working Party stated: "the information collected in the context of behavioural advertising [constitute personal data – and therefore information as an applicability criterion]."[50] Online behavioural advertising exemplifies a context in which artificial intelligence and machine learning processes produce novel information about individuals without human cognitive involvement.[51]

---

[49] In clarifying that the Data Retention Directive concerned the retention of personal data, the Court stated: "By requiring the retention of the data listed in art. 5(1) of Directive 2006/24 and by allowing the competent national authorities to access those data, Directive 2006/24, as the Advocate General has pointed out, in particular, in paragraphs 39 and 40 of his Opinion, derogates from the system of protection of the right to privacy established by Directives 95/46 and 2002/58 with regard to the processing of personal data in the electronic communications sector." *Digital Rights Ireland Ltd v Minister for Communications and others, and Kärntner Landesregierung,* Joined Cases C-293/12 and C-594/12 [2014] ECLI:EU:C:2014:238, para. 32 (hereinafter *Digital Rights Ireland*).

[50] Article 29 Working Party, "Opinion 2/2010 on Online Behavioural Advertising" (WP 171, 2010), p. 9.

[51] Contrary to Bygrave's interpretation of the ISO definition of information, which seems to put the focus exclusively on human cognition following the processing of data – as the carrier of information – cognition or knowledge representation are also an integral element of machine learning, so that computer agents can conduct automated reasoning. Bygrave, "Information Concepts in Law" *supra* n. 2, p. 91; Tim Berners-Lee, James Hendler and Ora Lassila, "The

The previous four sections mapped the substantive boundaries of the concept of information as an applicability criterion (role 1) in the GDPR. With this mapping complete, we now move on to perform the same process in relation to the concept of information as an object of regulation (role 2) in the GDPR.

9 Providing an Overview of the Background of Information as an Object of Regulation (Role 2)

The concept of information as an object of regulation in the GDPR was never explicitly recognised or elaborated by the legislator in the legislative process. There are thus no primary sources to consult to provide a background to the concept. However, the concept is implicit in the substantive principles in the GDPR and is reflected in the assumptions these embody. A look at the range and modalities of these provisions thus provides the basic material from which the concept can be mapped.[52]

Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities" *Scientific American* (New York, May 2001); Qihui Wu et al., "Cognitive Internet of Things : A New Paradigm Beyond Connection" (2014) 1(2) *IEEE Internet of Things Journal* 129-143. One can, therefore, point to some confusion around "the meaning of meaning" and of cognition. When a human sees the result of a data processing on the screen of the device, this will constitute information provided that the cognition process at the human level is successful. However, and regardless of that, cognition will have taken place at the level of the very processing itself. It is therefore important to distinguish between human and computer cognition, which do not overlap. See, for example: Frederik Zuiderveen Borgesius, "Personal data processing for behavioural targeting: which legal basis?" (2015) 5(3) *International Data Privacy Law* 163-176, p. 165. Indeed, from a more historical perspective, one could even consider the expanding scope of the notion of personal data from the perspective of human cognition. Earlier definitions of personal data, such as that adopted in the original French Data Protection Act, solely referred to "biographical information" (from the original French: "information nominative"). At this stage, the overlap between machine and human cognition was arguably total. However, with advances in computing, one can argue that the definition of personal data retained in the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, and subsequently Directive 95/46, made room for a non-human cognition aspect. See Jessica Eynard, *Les Données Personnelles: Quelle Définition Pour Un Régime de Protection Efficace?* (Paris, Michalon 2013), p. 11.

[52] This role for the concept of information logically relates to the first role for the concept of information – information as an applicability criterion. The fact the GDPR only applies to

The GDPR consists of three types of substantive provision relating directly to the handling and manipulation of information. First: legitimate processing provisions. All personal data processing must be legitimated under one of the grounds outlined in art. 6 – in relation to regular, non-sensitive personal data – or art. 9 – in relation to sensitive personal data.[53] In both cases, these legitimate grounds can, conceptually, be split into two groups: consent; and public interest justifications.[54] Second: data controller obligations. In principle, in all cases of personal data processing, the data controller must adhere to a set of obligations – centrally outlined in art. 5 – including, for example: the obligation to maintain personal data accurately; and the obligation to treat data confidentially.[55] Finally: data subject rights. In all cases of data processing, the data subject retains, in

substances which qualify as information provides the rationale for legislative consideration of information as an object of regulation. Accordingly, it should be noted that the concept of information as an object of regulation will not have been the only criterion the legislator will have had in mind when designing substantive provisions. Other art. 2(1) and 4(1) criteria will also have played a role. None of the GDPR's provisions on consent – for example art. 4(11) – elaborate what should happen in the case of a data subject's death. The reason is that the art. 4(1) applicability criterion of natural person excludes the deceased and thus the need to design provisions dealing with data protection and the deceased. See, for a discussion of the boundaries of the concept of natural person as well as the protection of post-mortem privacy under EU data protection law: Edina Harbinja, "Does EU data Protection Regime Protect Post-Mortem Privacy and what could be the Potential Alternatives?" (2013) 10(1) *SCRIPTed* 19-38, p. 27.

53 That personal data processing must always have a legitimation under art. 6 or art. 9 has been repeatedly confirmed in CJEU jurisprudence. See, for example: *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, para. 71 (hereinafter *Google Spain*). This case also references a long history of CJEU case law confirming the point. See, for example: *Worten – Equipamentos para o Lar SA v Autoridade para as Condições de Trabalho (ACT)*, Case C-342/12, [2013] ECLI:EU:C:2013:355, para. 33 (hereinafter *Worten*).

54 See for a discussion of the two types of legitimation ground: Omer Tene and Christopher Wolf, *The Draft EU General Data Protection Regulation: Costs and Paradoxes of Explicit Consent* (Future of Privacy Forum White Paper, 2013), p. 2, available at http://www.scribd.com/doc/121642539/The-Draft-EU-General-Data-Protection-RegulationCosts-and-Paradoxes-of-Explicit-Consent (accessed 03 May 2019).

55 That personal data processing must always adhere to the data protection principles outlined in art. 5 has also been repeatedly confirmed in CJEU jurisprudence. See, for example, *Google Spain*, *supra* n. 53, para. 71. This case also references a long history of CJEU case law confirming the point. See, for example, *Worten*, *supra* n. 53, para. 33.

principle, certain rights over their personal data including, for example: the right to withdraw consent; and the right to access personal data.

With few exceptions – notably the right to data portability in art. 20, the data protection impact assessment obligation in art. 35 and data breach notification obligations in arts. 33 and 34 – the substantive provisions outlined in the GDPR are not novel.[56] Most provisions were already present in some form in Directive 95/46. Most provisions present in Directive 95/46 were, in turn, inherited from provisions present in earlier EU Member States' data protection law and/or other international data protection instruments with European relevance. Indeed, as González Fuster observes, the core data controller obligations can be traced back to the first two international instruments with European relevance which emerged in the early 1980s: the Organisation for Economic Cooperation and Development's Guidelines on the Protection of Privacy and Transborder Flows of Personal Data (1980); and the Council of Europe's Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (1981).[57]

Each of the substantive provisions in the GDPR was designed to be flexible. Flexibility by design is necessary as a result of the omnibus nature of the GDPR and the need for further specification of provisions to account for sectoral processing differences.[58] Flexibility is also necessary in order for provisions to

---

[56] For a general discussion of the novelty of substantive provisions, see Christopher Kuner, "The European Commission's Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law" (2012) *Bloomberg BNA Privacy and Security Law Report* 1-15; Paul De Hert and Vagelis Papkonstantinou, "The new General Data Protection Regulation: Still a sound system for the protection of individuals?" (2016) 32(2) *Computer Law and Security Review* 179-194.

[57] Gloria González Fuster, *The Emergence of Personal Data Protection as a Fundamental Right of the EU* (Heidelberg: Springer, 2014), p. 84.

[58] See, for example, a discussion of EU data protection law as omnibus legislation in relation to the medical research context: Roberto Lattanzi, "Data Protection Principles and Research in the Biobanks Age" in Deborah Mascalzoni (ed.), *Ethics, Law and Governance of Biobanking* (Dordrecht: Springer, 2015), pp. 79-93, p. 85.

adapt to changing information processing technologies, the changing social contexts in which these technologies are used and the changing risks with which they are associated.[59] To provide more concrete interpretations of provisions, as necessary according to context, Data Protection Authorities (DPAs) – at both the EU (the EDPB) and Member State level (national DPAs) – are provided with broad interpretative and adaptive powers.[60] Flexibility of provisions has, however, significant limits. Applicable provisions cannot simply be disapplied. Nor can provisions be interpreted such that they disproportionately disapply other provisions; conflict with their own core purpose; conflict with the core aims of data protection law generally; or disproportionately impact other rights or legitimate interests engaged by data processing.[61]

Against this background, we now move to map the concept of information as an object of regulation against each of the three differentiating axes – outlined in section 3. As discussed above, this mapping cannot rely on primary sources. As an alternative, in relation to each axis, we consider which assumptions about the characteristics of information to be stored and manipulated must be present for the GDPR's substantive provisions to make sense.[62]

---

[59] See for a general discussion of the rationale and necessity of the flexibility of data protection principles in relation to changing technological and social consequences and changing risks to individuals' rights: Paul De Hert, "The Future of Privacy. Addressing Singularities to Identify Bright-Line Rules That Speak to Us" (2016) 2(4) *European Data Protection Law Review* 461-466.

[60] These powers even extend to the interpretation and adaptation of provisions considering novel technological and social challenges. For example: art. 70(1)(e) grants the EDPB the power to "[examine]…any question covering the application of this Regulation"; art. 58(3)(b) grants national DPAs broad discretionary powers to: "issue…opinions…on any issue related to the protection of personal data."

[61] *Supra* n. 30, pp. 403-405.

[62] We recognise that the methodology we use in this mapping process is somewhat unusual – particularly in a legal paper. However, we believe the methodology is both justified and unavoidable. We also recognise that, by our logic, an argument could be made for looking to map concepts of information along the three axes in relation to each different substantive principle. This possibility is a subject which should be followed up in further research.

10 Mapping the Concept of Information as an Object of Regulation in Terms of the Relationship between Information and Meaning (Role 2, Axis 1)

In the first instance, the concept of information as an object of regulation relates only to semantic information. Substantive provisions in the GDPR aim to describe the modalities of controller action in relation to the processing of individuals' personal data which may result in risks for those individuals. Such control mechanisms only make sense in relation to semantic information. Concepts of information where semantic content is not central, are thus irrelevant. Algorithmic information, for example – which defines the informational content of an object in terms of the bits of the smallest programme capable of calculating that object – is anathema to the concept of information as an object of regulation.[63] Bygrave's observation thus remains relevant: "Law is primarily concerned with regulating human relations; therein, the production and exchange of meaning play a key role."[64]

We can, however, identify one further boundary criterion. The concept of information as an object of regulation is limited to information of specific semantic content: highly structured information in the form of social facts. The effective function of multiple substantive provisions in the GDPR depends on the information being processed being in the form of social facts. The assumption is evident, in particular, in relation to provisions aimed at ensuring the transparency of data processing to the data subject – for example, consent provisions in art. 6(1)(a) and 9(2)(a), information obligations in arts. 13 and 14 and access rights in art. 15.[65] These provisions work on the basis of a one-off

[63] See, for example, Gregory Chaitin, *Algorithmic Information Theory* (Cambridge: CUP, 1987).

[64] Bygrave, "Information Concepts in Law" *supra* n. 2, p. 112.

[65] Indeed, these provisions have been highlighted by certain authors as constituting the core of the protection outlined by European data protection law. Deryck Beyleveld, "An Overview of

communication model, mandating the provision of a range of types of information about a processing operation to be provided to a data subject, such that the data subject is put in the position to understand the scope and consequences of processing.[66] Information provided should be accurate and relevant at the moment of information provision and remain accurate and relevant over the duration of processing.

Such a one-off communication model, however, does not necessarily function in relation to unstructured information – which requires further interpretation and structuring to produce socially relevant facts about a data subject. Two logical problems emerge. First, if the socially relevant factual content of information only surfaces via interpretation – during processing – this content only surfaces after the communication of information about the processing to the data subject. How can the data subject appreciate the consequences of processing, if they cannot be informed of the socially relevant factual content of the information about them which will eventually be processed?[67] Second, the socially relevant facts which can be extracted from the

Directive 95/46/EC in Relation to Medical Research" in Deryck Beyleveld, David Townend, Ségolène Rouillé-Mirza and Jessica Wright (eds.), *The Data Protection Directive and Medical Research Across Europe* (Aldershot: Ashgate, 2004), pp. 5-23, p. 11.

66 As the Article 29 Working Party stated: "A central consideration of the principle of transparency outlined in these provisions is that the data subject should be able to determine in advance what the scope and consequences of the processing entails and that they should not be taken by surprise at a later point about the ways in which their personal data has been used." Article 29 Working Party, *Guidelines on transparency under Regulation 2016/679* (WP260 rev.01, 2017 (revised 2018)), p. 7. In this regard, we would argue the information provided should allow the data subject to understand the scope and consequences of processing from a range of perspectives, including: (i) how the processing will impact the data subject's life – which types of actors are likely to make which significant judgments, in which contexts and with which likely outcomes for the data subject; (ii) the potential risks associated with the processing; and (iii) the range of options the subject has to actively influencing the processing – which rights the subject has in relation to processing and how these might be used.

67 As Albers observes: "data are not meaningful per se, but rather as 'potential information'." Marion Albers, "Realizing the Complexity of Data Protection" in Serge Gutwirth, Ronald Leenes and Paul De Hert (eds.), *Reloading Data Protection* (Dordrecht: Springer, 2014), pp. 213-235, p. 222.

analysis of unstructured information may change over time – as, for example, interpretative approaches advance. How can the data subject appreciate the consequences of processing, if the range of relevant facts which might be extracted from their information is liable to change?

We recognise a counter-argument might be put forward to the above position: many of the types of information required to be communicated under the GDPR's transparency provisions, which are relevant to allowing data subjects to understand the scope and consequences of processing, are unrelated to the degree to which information processed is already in the form of social facts. For example, transparency provisions contain a general obligation to communicate information concerning the purposes of the processing to the data subject – see, for example, arts. 13(1)(c), 14(1)(c) and 15(1)(a). Information on the purposes of processing is indeed vital for the data subject to understand the scope and consequences of processing. It is also true that information on the purposes of processing is technically independent of the degree to which processed information requires interpretation to produce social facts.

This counter-argument is superficially persuasive. The counter-argument, however, fails to recognise that: the factual content of information processed has, from the perspective of the consequences and risks of processing, a significant impact on how all other relevant information about processing is understood and evaluated. Consider, for example, the case of information concerning the purposes of processing about online behavioural advertising. The evaluation of the consequences and risks of such processing for a data subject's life will vary depending on the factual content of personal data being processed. Evaluation of consequence will differ, for example, depending on whether an advertiser processes information on a subject's shoe size, or whether they also process

information on a subject's sexuality.[68] Thus, the mere provision of information on the purposes of processing will not necessarily be sufficient to allow the data subject to understand the consequences and risks of processing.

11 Mapping the Concept of Information as an Object of Regulation in Terms of the Relationship between Information and Media (Role 2, Axis 2)

The concept of information as an object of regulation is limited to media which facilitate the easy and cost-effective reproduction and communication of information. In practice, this reduces the range of storage and transfer media that the concept encompasses artificial man-made media designed for easy reproduction and transfer of information – for example paper and digital media. This boundary criterion is an underlying assumption behind the effective function of several substantive provisions in the GDPR. Most significant amongst these provisions are access rights elaborated in art. 15 – particularly data subject rights to obtain a copy of personal data – and data portability rights outlined in art. 20 – in relation both to the right to obtain a copy of one's own personal data, and the right to have personal data transferred to another controller.

Arts. 15 and 20 function by permitting the data subject to easily and cheaply obtain, or have transferred, a copy of their personal data. The provisions constitute formal safeguards allowing data subjects transparency in relation to, and control over, the processing of their information, whilst not imposing

[68] See, for example, for a discussion of the specific consequences and risks to data subjects in the processing of sensitive types of personal data – including, according to art. 9(1), personal data concerning "sex life or sexual orientation" – in the context of online behavioural advertising: Information Commissioner's Office, *Update Report into adtech and real time bidding* (Report, 2019), p. 16, available at https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf (accessed 6 March 2020).

prohibitive costs or absurd modalities of action on data controllers.[69] Art. 15 requires that: "The controller shall provide a copy of the personal data undergoing processing." Art. 20 states: "The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format." The approach in these articles makes sense for artificial man-made data storage media designed for easy and cheap storage and transfer of information. Information stored and transferred in such media will tend to be accessible to data subjects and thus can facilitating transparency. The ease and cost of storage and transfer will then not impose undue burdens on data controllers.

Such an approach does not, however, function, when the media of storage and transfer are not artificial man-made media, but rather are naturally occurring media – for example human biological samples. Two logical problems appear. First, there is no guarantee that information stored and transferred in naturally occurring media will be readily accessible to data subjects. Thus, there is no guarantee that transfer of such media to data subjects will assist with transparency. Second, information stored and transferred in naturally occurring media are tendentially not amenable to cheap and easy reproduction or transfer. Thus, the imposition of reproduction and transfer obligations on data controllers is liable to impose prohibitive costs and absurd modalities of action. The reality of data controllers needing to engage in art. 15 or 20 obligations in relation to

[69] This is indicated by the legislator's express efforts to preclude the need for extreme resource deployment in the discharge of these rights. In relation to access rights: art. 15(3) recognises the right of the data controller to avoid such expense in levying charges on the data subject for the provision of any more than one copy of their personal data: "For any further copies requested by the data subject, the controller may charge a reasonable fee based on administrative costs." In relation to portability rights: Recital 68 relieves data controllers from the need to adopt special systems to ensure common formats across data controller systems: "The data subject's right to transmit or receive personal data concerning him or her should not create an obligation for the controllers to adopt or maintain processing systems which are technically compatible." See also Article 29 Working Party, *Guidelines on the right to data portability* (WP 242, 2016), pp. 13-14.

naturally occurring media may seem unlikely. In section 15, however, we will provide concrete examples.

12 Mapping the Concept of Information as an Object of Regulation in Terms of the Relationship between Information and Human Cognition (Role 2, Axis 3)

In the first instance, the concept of information as an object of regulation requires human cognitive involvement in establishing the processing context. Virtually all substantive principles in the GDPR are based on the presumption of human cognitive influence over the processing context. All core data controller obligations outlined in art. 5, for example, require that social considerations to have taken place in establishing the modalities of a processing context, for which human cognition is a prerequisite. For instance, discharge of the art. 5(1)(f) obligation concerning the confidential collection and storage of information requires human cognitive involvement in at least two ways. First, human cognitive involvement is required in defining which relational boundaries should be considered as demarking relationships of confidentiality.[70] Second, human cognitive involvement is then required to determine the degree to which technical and organisational approaches are necessary to maintain the integrity of these boundaries.[71]

We can, however, also identify one further boundary criterion: the concept of information as an object of regulation requires human cognition to be capable of perceiving, and comprehending, the content of information being processed – even if perception never, in fact, takes place. This requirement is implicit in the

[70] See, for example, the social calculations involved in confidentiality requirements in UK medical law: Nick Nicholas, "Risk management: Confidentiality, disclosure and access to medical records" (2007) 9 *The Obstetrician and Gynaecologist* 257-263, p. 258.

[71] Frederik Zuiderveen Borgesius and Dara Hallinan, "Article 5" in Franziska Boehm and Mark Cole (eds.), *GDPR Commentary* (Cheltenham: Elgar, Forthcoming 2020).

effective function of a broad range of substantive provisions in the GDPR.[72] Certain provisions require the possibility for human perception and understanding of information to ensure adequate control measures are, or have been, implemented. For example, effective discharge of the art. 5(1)(d) accuracy obligation requires the possibility for human perception and understanding of information processed to ensure information are accurate and up to date.[73] Other provisions require the possibility for human perception and understanding of information to ensure suitable manipulation of information occurs, or has occurred. For example, art. 17 erasure requirements require individuals to implement, and confirm, information erasure. Yet other provisions require the possibility for human perception and understanding of information to ensure adequate external communication of information concerning the details and consequences of processing operation can occur. For example, data subject transparency rights stipulated under art. 14 require a controller to communicate to the data subject "the categories of personal data [being processed]."

Despite the above observations, it should be highlighted that the concept of information as an object of regulation does not foresee the need for human cognition in relation to the creation or perception of each and every specific element of information processed. None of the substantive principles in the GDPR function on the basis that information must have been created or perceived by a human. In fact, there are provisions in the GDPR which relate solely to processing contexts in which information is created and processed with no direct human cognitive involvement at all. Arts. 21 and 22, for example, relate to instances in which automated profiling and decision making – situations

[72] Indeed, the only provisions for which this is not true are those related to activities to be carried out prior to processing – for example provisions relating to the obligation to conduct a data protection impact assessment in art. 35.

[73] Jiahong Chen, "The Dangers of Accuracy: Exploring the Other Side of the Data Quality Principle" (2018) 4(1) *European Data Protection Law Review* 36-52, pp. 37-38.

including artificial intelligence and machine learning – are in play.[74] These articles do not serve to diminish the applicability of other substantive principles in the GDPR, but simply offer supplemental protection when no human is engaged in the creation or perception of information.

The previous eight sections mapped the substantive boundaries of the concept of information as an applicability criterion and as an object of regulation. Considering the results of these mapping processes, we now move to compare the boundaries of the concepts occupying these two roles to highlight that the substantive boundaries of the concepts do not converge on the same substantive phenomenon.

13 The Two Concepts of Information Relate to Different Substantive Phenomena

A comparison of the two concepts of information reveals multiple points of difference. We distinguish two key types of difference: first, differences in degrees of flexibility; second, differences in substantive boundaries. On the back of the consideration of points of difference between concepts, we venture a future prediction as to how differences between concepts will develop.

In terms of the flexibility of the two concepts: via a comparison of the backgrounds of the concepts, it is evident the concept of information as an applicability criterion is considerably more flexible than the concept of information as an object of regulation. The flexibility of the concept of information as an applicability criterion is extreme.[75] Eventually, this concept of

[74] See Antoni Roig, "Safeguards for the right not to be subject to a decision based solely on automated processing (art. 22 GDPR)" (2017) 8(3) *European Journal of Law and Technology*, p. 2.

[75] This recognition has a further consequence which deserves more extensive discussion elsewhere: this concept of information may map to phenomena not corresponding to traditional understandings of information at all. Certain effects with which EU data protection law is concerned relate to presumptions of information processing. One example would be chilling effects. The CJEU observed the relevance of chilling effects in relation to information

information is normatively defined by its role in "turning on" the system of protection offered under the GDPR. The concept thus potentially becomes relevant whenever risks for the individual in relation to rights in the information society are identifiable.[76] The flexibility of the concept of information as an object of regulation, however, is limited. The flexibility of this second concept of information is tied to the flexibility of the GDPR's substantive provisions. These are indeed imbued with a degree of flexibility. These provisions consist, however, of concepts and relationships with defined boundaries in both natural language and law. These boundaries cannot be ignored and thus serve as unavoidable restrictions on the elasticity of provisions.

In terms of the substantive boundaries of the two concepts: variation is identifiable along each of the three axes making up the structured conceptual framework differentiating concepts of information. In relation to axis 1: information as an applicability criterion encompasses all semantic information, while information as an object of regulation relates only to semantic information in the form of social facts. In relation to axis 2: information as an applicability criterion encompasses all information storage and transfer media, while information as an object of regulation relates only to media which facilitate

processing in the *Digital Rights Ireland* case. *Digital Rights Ireland*, *supra* n. 49, para. 28. There may be cases in which no information processing actually occurs and yet chilling effects risks related to information processing are still relevant. For example, dummy camera systems: no information processing occurs, but the systems engage chilling effects risks concerned with the presumptions of information processing. In this case, the phenomenon constituent of the relationship between the data subject and the data controller is not informational, but doxastic. It has thus been suggested that the concept of information, via teleological interpretation, could also extend to doxastic relationships. See, for a discussion of this possibility, Dara Hallinan, "Data Protection without Data: Could Data Protection Law Apply without Personal Data Being Processed?" (2019) 5(1) *European Data Protection law Review* 293-299.

[76] Recall the CJEU observation in the *Nowak* case – see section 5: "The use of the expression 'any information' in the definition of the concept of 'personal data', within Article 2(a) of Directive 95/46, reflects the aim of the EU legislature to assign a wide scope to that concept…potentially encompasses all kinds of information…provided that it 'relates' to the data subject." *Nowak*, *supra* n. 21, para. 34.

frictionless copying and transfer of information – artificial man-made media. In relation to axis 3: information as an applicability criterion requires human involvement in setting the processing context, while information as an object of regulation additionally requires human cognitive ability to perceive and understand information.

Moving forward, as a result of differences in both flexibility and substance, we predict the substantive gap between the two concepts will become more pronounced over time. On the one hand, the scope of the concept of information as an applicability criterion will likely expand. As Purtova argues, social relationships are become increasingly informationally mediated and an increasing range of objects and processes are being perceived in terms of information. The range of social interactions, objects and processes which are capable of giving rise to threats to individuals' rights in information thus grows accordingly.[77] The scope of data protection law under the GDPR – and therefore the concept of information as an applicability criterion – will thus need to expand in response to this phenomenon. Data protection, after all, is specifically tasked with protecting individuals' rights engaged by the collection and processing of their information.[78] On the other hand, the concept of information as an object of regulation will likely remain comparatively static. This seems likely given the inherent limitations in the flexibility of the substantive principles of data protection law, as well as given the lack of legislative recognition, or effort to update, presumptions supporting the concept to date.

---

[77] See, for example, 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* 40-81, p. 43.

[78] This is not to say there are no other areas of law which play a role in the protection of individuals' rights in information – the rights to privacy and to freedom of information and duties of confidentiality, for example. Our thanks to anonymous reviewer 2 for pointing this out.

This section summarised the differences – in terms of both flexibility and substantive boundaries – between the two concepts of information in data protection law under the GDPR. Because of these differences, the GDPR will apply to types of information for which its substantial principles were not designed. In light of this assertion, we thus move to outline our third thesis: *The substantive differences between the two concepts of information are significant for the efficacy of the GDPR as an instrument of information law.* To elaborate the thesis, we will provide examples of problems with the GDPR in relation to contemporary data processing phenomena which can be linked – at least partly – to differences between the two concepts of information. We provide one problematic example along each of the three axes comprising the structured conceptual framework differentiating concepts of information.[79]

14 The Consequences of Differences in Concepts of Information in the GDPR: Problems with differences in concepts relating to semantic meaning (axis 1)

The processing of genomic sequence information provides an example of a phenomenon in which problems emerge as a result of differences in concepts of information relating to the semantic meaning of information.

[79] We would like to highlight that we are in no way suggesting that issues with the lack of consideration, or clarity in the substantive definition, of the concept of information are definitive for the range of possible problems with the GDPR. Nor do we wish to suggest that problems which may be framed in terms of differences in information cannot also be framed – perhaps much more fruitfully – in other ways. Rather, we aim to highlight that the different conceptualisations of information active in the GDPR play some role in the efficacy of the GDPR as an instrument of law. With this observation, we hope to spark further research into the conceptualisation, and the significance of the conceptualisation, of information in EU data protection law.

Genomic sequence information – as a form of semantic information – conforms to the salient criteria of information as an applicability criterion.[80] Genomic sequence information does not, however, conform to the salient criteria of the concept of information as an object of regulation, as it is not in the form of social facts, but rather is in the form of unstructured information requiring further interpretation to be turned into social facts. The range of possible interpretations which may be applied to genomic sequence information at any given time – and thus the range of social facts which might be produced from genomic sequence information at any given time – depends on the state of genetic science at the time. In turn, the future development of genetic science is highly unpredictable.[81] As Pontin has observed of genetic science's efforts to get to grips with the function and content of the human genome: "no one will contest that the genome has turned out to be bafflingly complex."[82]

The logical problems highlighted in section 10 concerning the inability of the GDPR's transparency provisions' – for example those in arts. 13, 14 and 15 – one-off communication approach in dealing with unstructured information thus become reality in relation to processing involving genomic sequence information. How, for example, should a data controller processing genomic sequence information, as obliged by art. 14, give a data subject a useful list of "categories of personal data" to be processed which will remain accurate over time. The controller could, at best tell the data subject that their genomic sequence

---

80 This assertion has been repeatedly confirmed in jurisprudence. The Article 29 Working Party for example, have stated: "genetic data are doubtlessly 'personal data'." Article 29 Working Party, *Working Document on the processing of personal data relating to health in electronic health records (EHR)* (WP 131, 2007), p. 7.

81 It seems all but inevitable that further scientific advance will follow, leading to the ability to extract yet more facts about individuals from the genome sequence. For a discussion see Chris Tyler-Smith et al., "Where Next for Genetics and Genomics?" (2015) 13(7) *PLOS Biology*.

82 Jason Pontin, "A Decade of Genomics: On the 10th anniversary of the Human Genome Project, we ask: where are the therapies?" (*MIT Technology Review*, 21 December 2010), available at https://www.technologyreview.com/s/422130/a-decade-of-genomics/ (accessed 10 March 2020).

information will be processed and give an accompanying rundown of the types of social facts which can, at the moment of communication, be extracted from the sequence. Such a provision of information, however, would do nothing to address the fact that new types of socially relevant facts will become extractable from the sequence as genetic science advances.[83]

We recognise a counter-argument might be put forward suggesting that the severity of this issue will, in practise, be mitigated by common knowledge concerning the possibility to interpret information to produce social facts. From this perspective, common knowledge provides an epistemic framework, generally available to data subjects, which renders the need for anything more than a one-off communication moot. This would be a strong argument should detailed common knowledge on the interpretability of the genome sequence really be prevalent among EU citizens. This, however, seems unlikely to be the case. In this regard, Lanie et. al., in summarising their survey of public understanding of genetics and genomics, state: "this study provides…evidence…demonstrating that misconceptions about genetic science are not infrequent in the general public, and suggests the need for improved genetic literacy and understanding."[84]

15 The Consequences of Differences in Concepts of Information in the GDPR: Problems with differences in concepts relating to media (axis 2)

The processing of biological samples provides an example of a phenomenon in

[83] Indeed, in this vein, there have already been discussions of the inadequacy of one-off communications models in relation to informed consent in the processing of genomic sequence information in genomic research. See Christine Grady et al., "Broad Consent For Research With Biological Samples: Workshop Conclusions" (2015) 15(9) *American Journal of Bioethics* 34-42, p. 43.

[84] Angela Lanie et. al., "Exploring the Public Understanding of Basic Genetic Concepts" (2004) 13(4) *Journal of Genetic Counselling* 305-320, p. 318.

which problems emerge as a result of differences in concepts of information relating to the medium of information storage and transfer.

Biological samples constitute an information storage and transfer medium which fulfil the salient criteria of information as an applicability criterion. As naturally occurring media – rather than artificial man-made media – however, they do not conform to the salient criteria of the concept of information as an object of regulation. The logical problems highlighted in section 11 concerning the application of the GDPR's data transfer provisions – for example those in art. 15 and in art. 20 – to processing involving naturally occurring media thus become reality in relation to the processing of biological samples. The provision of a copy of a biological sample to a data subject – for example to a genomic research subject – will not serve to allow the subject to better understand the processing being conducted. The subject is highly unlikely to have the means to easily access the information in the sample and even if they did, this would do little to assist them in understanding the processing taking place. At the same time, the need to replicate and transfer the sample may impose large costs on a data controller.[85] For example, a copying process, as Mason observes, would require a "disproportionate cost" in terms of producing an immortal cell-line from the sample – through a process such as a polymerase chain reaction.[86]

---

[85] *Supra* n. 30, pp. 380-385.

[86] "The data subject could be given a sample of 'relevant' genetic material amplified by polymerase chain reaction (though at disproportionate cost!)." Neil Mason, "The medium and the message: tissue samples, genetic information and data protection legislation" in Heather Widdows and Caroline Mullen (eds*.), The Governance of Genetic Information: Who Decides?* (Cambridge: CUP, 2009) pp. 15-36, p. 29. A transfer process may also require specially designed transport facilities to effectively move the biological sample – such as refrigerated vehicles. Kunkel et. al., for example, observe that transport of certain biological samples would require the "maintenance of ultra-low conditions at all stages during transport…obtained with high-quality packaging and dry ice or liquid nitrogen in quantities sufficient to last during unforeseen delivery delays." Eric Kunkel, Rolf Ehrhardt, "Frozen Assets – An Expert Guide to Biobanking" (*Select Science*, 23 December 2014), available at http://www.selectscience.net/editorial-articles/frozen-assets--an-expert-guideto-biobanking/?artID=35743. Last consulted: 20.04.2018 (accessed 9 March 2019).

A counter-argument could be put forward that discussion of this problem is based on the fallacious assumption that the handling of biological samples will fall under the scope of the GDPR. This argument is built on the fact that art. 2(1) clarifies that the GDPR only applies to personal data which is "processed wholly or partly by automated means… and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing." On the back of art. 2(1), the argument then asserts that the handling of biological samples will not constitute processing "either wholly or partly by automatic means" or processing "which form[s] part of a filing system." It is certainly true that these art. 2(1) limitations will exclude certain activities involving the handling and use of biological samples from falling within the scope of the GDPR – for example, the use of biological samples in transplantations.

Yet, we would suggest that the counter-argument fails to consider the breadth of the relevant definitions in the GDPR, and therefore cannot be accepted. In art. 4(2), the GDPR recognises the concept of processing to encompass: "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration." Under such a definition, there are contexts in which activities involving the handling of biological samples will constitute processing. The biobanking context, for example, involves the methodical collection, recording and organisation of samples. Indeed, key definitions of biobanking – such as that provided by the National Health and Medical Research Council – explicitly highlight the activity as being defined by the collection and organisation of biological samples in a filing system.[87] Once it has been established that biological

[87] See, for example, the definition for biobanking provided in: National Health and Medical Research Council, *Biobanks Information Paper* (E110, 2010), p. 7, available at

samples can be processed, it is then a short step to recognise that biological samples can be automatically processed, or manually processed as part of a filing system – activities which are medium independent.

16 The Consequences of Differences in Concepts of Information in the GDPR: Problems with differences in concepts relating to human cognition (axis 3)

The processing of personal data in neural networks provides an example of a phenomenon in which problems emerge as a result of differences in concepts of information relating to human cognition.

The processing of personal data in neural networks corresponds to the salient qualities of the concept of information as an applicability criterion. Personal data in neural networks need not, however, conform to the salient qualities of information as an object of regulation, as these networks may not permit human cognitive perception or understanding of all the information they process. As Kamarinou et. al. observe, whilst certain types of algorithmic processing – for example decision trees – allow human cognitive perception and understanding of information processed: "The situation may be very different in relation to neural network-type algorithms, such as deep learning algorithms…the conclusions reached by neural networks are 'non-deductive and thus cannot be legitimated by a deductive explanation of the impact various factors at the input stage have on the ultimate outcome'."[88] The logical problems highlighted in section 12 concerning the need, in most provisions of the GDPR,

https://www.nhmrc.gov.au/about-us/publications/biobanks-information-paper (accessed 20 February 2020).

88 Dimitra Kamarinou, Christopher Millard, and Jatinder Singh, "Machine Learning with Personal Data" (Queen Mary University of London, School of Law Legal Studies Research Paper 247/2016, 2016), p. 19. The authors cite David Warner Jr., "A Neural Network-based Law Machine: The Problem of Legitimacy" (1993) 2(2) *Law, Computers & Artificial Intelligence* 135-147, p. 138.

for human cognitive ability to perceive and understand information thus become reality in processing involving neural networks.

In this regard: how, in any processing context involving complex and opaque neural networks, could a data controller be certain to have implemented suitable and adequate control mechanisms, to have made sure correct information manipulation has taken place or to have ensured that external communication of information has occurred, when they cannot perceive or understand the information being processed? In relation to art. 5(1)(d) obligations that information be held accurately, for example, Goodman et. al. highlight the difficulty in effectively evaluating information processed in a neural network: "what hope is there of explaining the weights learned in a multilayer neural net with a complex architecture."[89] Equally, in relation to art. 17 erasure requirements, Fosch Villaronga et. al. highlight the general difficulty in deleting data from artificial intelligence systems.[90] This difficulty is magnified manifold when the forms and functions of information within the system are opaque to those who must perform the deletion operation.

We recognise that the issues raised by neural networks – as well as other complex artificial intelligence and machine learning processing – in terms of the effective function of pertinent provisions of the GDPR have already been framed, and discussed, at length. This is particularly the case in relation to discussions of algorithmic transparency. Relevant authors in this regard include, amongst many others, Binns, Brkan, Kaminsky, Mendoza et. al., Selbst et. al., Wachter et.

---

89 Bryce Goodman and Seth Flaxman, "EU regulations on algorithmic decision-making and a 'right to explanation'", (2016) ICML Workshop on Human Interpretability in Machine Learning, p. 29, available at http://metromemetics.net/wp-content/uploads/2016/07/1606.08813v1.pdf (accessed 10 March 2020). The authors are, in this paper, discussing the right to an explanation under the GDPR in relation to artificial intelligence. The observation, however, is also relevant in this content.

90 Eduard Fosch Villaronga, Peter Kieseberg, Tiffany Li, "Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten" (2018) 34 *Computer Law and Security Review* 304-313, pp. 308-309.

al. etc.[91] Each of these authors has considered the context, function, input and output of personal data processed in artificial intelligence systems in relation to the effective function of substantive provisions in the GDPR. We understand questions might thus be raised as to whether it makes sense to consider the problems neural networks pose to the GDPR in terms of distinctions between concepts of information: what would such an approach bring?

In this regard, we do not see that a consideration of differences in concepts of information in the GDPR in relation to problems posed by neural networks poses any conceptual challenge to current discussions on algorithmic transparency. We do, however, believe the approach offers a new perspective within which to frame algorithmic transparency discussions. As highlighted by Gellert, algorithmic transparency discussions have hitherto focused overwhelmingly on the function of algorithms.[92] In these discussions, the concept of information has been largely ignored. Further research will be necessary, however, to conclude whether considering these issues from the perspective of differences between concepts of information – as opposed to algorithmic function – will bring fresh insight to discussions.

---

[91] Reuben Binns, "Algorithmic Accountability and Public Reason" (2018) 31 *Philosophy and Technology* 543-556; Maja Brkan, "Do Algorithms Rule the World? Algorithmic Decision-Making and Data Protection in the Framework of the GDPR and Beyond" (2019) 27(2) *International Journal of Law and Information Technology* 91-121; Margot Kaminski, "The Right To Explanation, Explained" (2019) 34 *Berkeley Technology Law Journal* 189-218; Isak Mendoza and Lee Bygrave, "The Right Not to Be Subject to Automated Decisions Based on Profiling" in Tatiana-Eleni Synodinou, Philippe Jougleux, Christiana Markou, and Thalia Prastitou (eds.), *EU Internet Law: Regulation and Enforcement* (Cham: Springer, 2017), pp. 77-98; Andrew Selbst and Julia Powles, "Meaningful information and the right to explanation" (2017) 7(4) *International Data Privacy Law* 233–242; Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation" (2017) 7(2) *International Data Privacy Law* 76-99.

[92] As Gellert also argues, data protection law currently regulates machine learning by bypassing the crucial aspect of learning and the informational concepts this presupposes. Gellert, "Data Protection and Notions of Information", *supra* n. 2, p. 20.

This section showed that divergence between the two concepts of information in the GDPR leads to problems for the efficacy of the GDPR as an instrument of information law. The next section concludes the paper by looking at the legal avenues through which such problems might be addressed.

17 Legal Avenues for Addressing Problems Relating to the Divergence of Concepts of Information

There are several legal approaches via which issues relating to the disparity between the two concepts of information might be addressed within the structure of the GDPR. Two are particularly important: DPA interpretation and adaptation; and Member State derogatory legislation.[93] Each of these approaches, however, has limitations. Eventually, legislative intervention and correction may be required.

Certain problems emerging from the divergence in the two concepts of information might be addressed via the GDPR's internal adaptive mechanisms. The GDPR foresees the possibility for the adaptation of substantive principles through national DPA – art. 57 – and EDPB – art. 70 – guidance. Such interpretation could help align the two concepts of information to address concrete problems. For example, EDPB guidance could clarify the applicability of art. 15 to biological samples such that the article no longer imposes absurd requirements on controllers. These mechanisms, however, are limited in their capacity to provide comprehensive solutions to divergences. The mechanisms

[93] We also recognise that Courts – at national and EU level – can also function as important mechanisms for the resolution of issues created by divergences between concepts of information in the GDPR. We refrain from discussing their role in this regard, however, owing to the fact that the likelihood of specific cases landing before national or EU courts dealing with these specific issues is hard to predict. As a result, it is hard to assert that Courts will be in the position to regularly act as a mechanism for the resolution of issues. For example, only very few cases dealing with issues concerning biological samples as information have ever come before courts in the EU.

are limited to the adaptation of principles already present in the GDPR. They thus have little capacity to overrule principles in the GDPR when these make no sense in relation to certain modalities of information. Nor have they the capacity to introduce new principles which may be necessary to provide supplemental protection in relation to modalities of information inadequately protected under current provisions.[94]

Certain problems could also be addressed via Member States making use of derogation possibilities to void principles of the GDPR. For example, art. 9(4) offers EU Member States the possibility to derogate from the GDPR in defining supplemental principles applicable to the processing of sensitive personal data while art. 89 permits Member States to derogate from certain substantive principles in the GDPR in relation to scientific research. Member State derogation could also help align the two concepts of information to address specific concrete problems. This approach, too, however, is subject to limitations and thus cannot provide a comprehensive solution. From a substantive perspective, there are limits to the principles from which EU Member States can derogate, the circumstances under which derogation is possible and the degree to which derogation is possible.[95] In turn, from a legal-structural perspective, any EU Member State use of derogation possibilities would disrupt the harmonious applicability of the GDPR across Europe.

Eventually, given that the GDPR's internal mechanisms are limited in their ability to resolve problems related to discrepancies between concepts of information, a more drastic solution comes into view: legislative intervention.

---

[94] There is no discussion of the ability to add to, or to exclude, the applicability of substantive provisions of EU data protection law in the outline of the powers held by DPAs or by the EDPB in art. 57 and art. 70, respectively, of the GDPR.

[95] See, for example, the possibilities, and discussion of uncertainties, in relation to Member State derogations under art. 89: Stephan Pötters, "Artikel 89" in Peter Gola (ed.), *DS-GVO DatenschutzGrundverordnung VO (EU) 2016/679 Kommentar* (2nd ed.) (Munich: Beck 2017), pp. 990-999.

The legislator would be ideally placed to introduce different strands in data protection law tailored for dealing with the different modalities of information to which the GDPR must apply. Indeed, long term, as discrepancies between the two concepts of information likely increase, and problems stemming from these discrepancies become more pronounced, legislative intervention may prove the only feasible way forward. We would observe however, that the legislator has barely, thus far, recognised the possibility that different modalities of information exist, nor that such differences may require tailored regulatory responses. In the legislative process leading up to the GDPR, for example, the idea of different modalities of information – as opposed to different actors, sectors and technologies – was scarcely thematised at all.[96]

Thus, there are possible approaches available to address the divergence in concepts of information internal to the GDPR. However, these internal mechanisms have limits. Eventually, to provide a comprehensive solution, the legislator may need to step in and recognise the existence of, and design bespoke standards of substantive protection in relation to, different modalities of information.

18 Conclusion

The concept of information plays two distinct roles in the GDPR. First, the concept functions as one of the GDPR's applicability criteria – as outlined in art. 4(1): information as an applicability criterion. Second, the concept refers to a substance around which the substantive provisions of the GDPR have been

[96] See, for example, the document initiating the reform process leading to the GDPR and its lack of reference to the various possible modalities of information as an issue to be addressed: European Commission, *A comprehensive approach on personal data protection in the European Union* (COM(2010) 609 final, 2010), pp. 2-5, available at https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0609:FIN:EN:PDF (accessed 6 March 2020).

designed, and in relation to which the substantive provisions in the GDPR are intended to act: information as an object of regulation. Significantly – albeit somewhat counterintuitively – the substantive boundaries of the concepts of information occupying these two roles do not converge on the same substantive phenomenon.

The concept of information as an applicability criterion is highly flexible and relates to all semantic information, stored and transferred in any media and processed in any processing context established with human cognitive involvement. The boundaries of the concept of information as an object of regulation are more concrete and encompass only semantic information in the form of social facts, stored and transferred in media which facilitate easy and cost-effective replication and transfer of information – i.e. artificial man-made media – processed in a context set via human cognitive involvement and amenable to human cognitive perception and understanding.

Differences between the two concepts are not simply academic curiosities. The divergence is a causal factor in a range of concrete problems with the efficacy of the GDPR as an instrument of information law. The broader scope of information as an applicability criterion means that the GDPR will apply to types of information for which its substantial principles were not designed. For example, the concept of information as an applicability criterion extends to information stored in the naturally occurring media biological samples. Yet, the concept of information as an object of regulation is limited to certain types of artificial man-made media. Consequently, when the GDPR is applied to biological samples, problems ensue. These include the potential imposition of absurd obligations on data controllers – such as obligations to copy and transfer biological samples to data subjects at disproportionate cost.

Moving forward, discrepancies between the two concepts of information might be addressed in piecemeal fashion, via specific solutions targeted at

specific problems. These solutions might be provided via EDPB and national DPA guidance providing interpretations of the GDPR tailored to address problems. These solutions may also be delivered via Member States using derogatory powers to disapply problematic provisions in the GDPR. In order to address the issue in a more comprehensive manner, however, direct legislative attention may be necessary. Indeed, if the GDPR is to continue to play a role as a key instrument of protection of individuals rights in modern information societies, then explicit legislative differentiation between modalities of information may eventually be necessary.

19 Acknowledgements

Raphaël Gellert has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (INFO-LEG project, grant agreement No 716971).