UiO : Faculty of Law
University of Oslo

Revisiting the Definition of Personal Data in the EU Data Protection Regime

Candidate number: 9011

Submission deadline: 15.05.2016

Number of words: 17987

Table of Contents

Abbreviations

A29WP	Article 29 Working Party
BRAIN	Brain Research through Advancing Innovative Neuroethologies
CJEU	Court of Justice of the European Union
DNA	Deoxyribonucleic acid
DPA	Data Protection Authorities
DPD	Data Protection Directive
ECHR	European Convention on Human Rights
ECtHR	European Court of Human Rights
EU	the European Union
GDPR	General Data Protection Regulation
HEP	Human Enhancement Projects
HGP	Human Genome Project
ICO	Information Commissioners Office (UK)
IP	Internet Protocol
ITC	Information and Communications Technology
PII	Personally Identifiable Information
UCLA	University of California at Los Angeles
UK	the United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organization

Chapter 1

Background and General Introduction

1 Introducing the Problems

There were numerous reasons for enacting the first data protection laws in the 1970s. Among the important factors was greater dissemination, use, and re-use of personal data across organizational boundaries which was facilitated by new technology in the form of electronic data processing which, in turn, engendered public fear of disempowerment, loss of control over technology and automation of societal processes.[1] In addition to rapidly increasing capacity to store data, computers permitted information to be searched and organized by multiple attributes, rather than through a single index (for example, first and last name only). This capacity changed the way information could be linked to an individual[2] which led to data protection laws focused on protecting "personal data" in the EU and "Personally Identifiable Information (PII)" in the United States of America.[3] The definitions of these key concepts delimit the scope of application of data protection laws. One of the major changes in the EU after the adoption of the directive has been the recognition of data protection as a fundamental right in itself, independent from the right to respect for private life.[4]

[1] Bygrave (2014), p. 8-15. See also, A29WP, Opinion 4/2007, p.5. Recital 4 in the preamble to the DPD makes a similar assertion

[2] Schwartz & Solove (2011), p. 1820

[3] The U.S., however, lacks a comprehensive set of data protection rules as is available in Europe and relies instead on sector specific rules. (See, Bygrave (2014), p. 110-12)

[4] See Article 16 of the Treaty of the Functioning of the European Union and Article 8 of the Charter of Fundamental Rights of the European Union

Today, more than 40 years after the enactment of the earliest data protection laws[5] and two decades after the EU Data Protection Directive was adopted, the technological landscape has dramatically changed. Computer power continues to grow[6] with additional capacities and data processing capabilities. The growth in computer power has aided a significant transformation in many fields of study including molecular biology and nano-technology. Consequently, there is strong criticism on sustainability of the EU data privacy laws' definition of personal data.[7] According to this definition personal data is, in essence, *information* which is capable of *identifying* living human data subjects. I call the two terms '*information'* and '*identifying'* the two building blocks of the definition, and it is the conceptual predispositions behind these terms that I seek to challenge by building on previous works.

The propriety of defining data as exclusively 'informational' is being put to test as advancements in bio-technology and Information and Communications Technology (ICT) continue to blur the distinction between the human biological materials and the information derived from them.[8]

The 'identifiability' criterion is also flawed as it continues to exclude so called 'anonymous' data/information from the scope of data protection regimes despite easy re-

[5] The first national data protection law was enacted by Sweden in 1973 (Sweden's Data Act) which is repealed and replaced by Personal Data Act of 1998; the very first legislation directly dealing with data protection was Hessian Data Protection Act enacted in 1970. (See, Bygrave (2014), p. 100)

[6] According to the notorious '*Moore's Law'* (an observation named after Gordon E. Moore of Intel) computer power (i.e. transistor count on an integrated circuit) continues to double every two years at least for another decade.

[7] Article 2(a) of the DPD reads: 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified , directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity

[8] See Bygrave (2010); Bygrave (2015); Taylor (2012), Chapter 7; Ploeg (2007)

identification.[9] The concept is also not suited to address privacy and related concerns when group interests are implicated.[10]

2 Objectives

Generally, this study's purpose is to analyze and highlight the problems and resulting difficulties in using and applying the current definition of personal data in the EU data protection regime. Specifically, the study seeks to:

- Clarify whether bio-materials constitute personal data under the existing legal framework (EU Data Protection Directive and the different versions of the proposed General Data Protection Regulation (GDPR))[11]
- Analyze the viability of maintaining a conceptual distinction between data and information
- Examine the pragmatic reasons for treating biological materials as data
- Scrutinize the practical effects of and problems associated with considering biological materials to be personal data
- Evaluate the concept of 'identifiability' and the major problems arising from the use of the criterion
- Evaluate the extent to which the 'identifiability' criterion serves new forms of group interests

3 Methods and Limitations

3.1 Methods

Generally speaking, the nature of a research question determines the method that should be employed.[12] This study employs both *descriptive and normative* legal research methods to critically examine what the law is, and what it ought to be.

[9] See Ohm (2010); Schwartz & Solove (2011)

[10] See, for instance, Taylor(2012), Chapter 5; Mantelero (2016) p. 256-271

[1111] I analyzed various version of the regulation in because different changes that occurred in the process are relevant to the discussion.

The *descriptive* component explains and clarifies the EU rules on this area (*de lege lata*). The Data Protection Directive (DPD) and the General Data Protection Regulation serve as a main focus with reference to national laws when relevant. After the valid scope and content of the existing regime is identified, the *normative* aspect is used to assess the adequacy of the existing framework and the new data protection regulation. (*de lege ferenda*)

Due to the evolving nature of the regulatory landscape in the area, the research utilizes not only hard laws in effect but also a range of other initiatives such as EU Commission proposals, readings of the European Parliament and the Council. Various explanatory memoranda and official commentaries are scrutinized as well.

Moreover, the interdisciplinary nature of the research demands analysis of wide range of materials collected not only from legal literature, but also from computer science, social philosophy, biotechnology, and molecular biology. Only in this way do we gain a better understanding of the operation of a given technology, how it affects our socio-economic relations and, subsequently, how it should be regulated by law.

3.2 Limitations and Challenges

One of the challenges in approaching this research is its *interdisciplinary nature*. The study raises multiple questions about the propriety and implications of regulating, biological material by means of data protection law. Further, it delves in to the usefulness of 'identifiability' as a mechanism to properly address privacy concerns.

These tasks could seem daunting tasks for lawyers, particularly because they involve a number of complex technological developments in bio-technology, nano-technology, and information and communications technology. The study, therefore, takes in to account such difficulty and consults the necessary literature from the fields of Bio- technology, Social Philosophy and, Computer and Communications Technology.

[12] Schrama (2011), p. 148

Another major challenge comes from the changing nature of the regulatory landscape under study.[13] The study was carried out while the regulatory framework itself was evolving. Many of the proposals and rules analyzed have undergone changes while the research was in progress. This challenge necessitates a constant inspection of the rules and processes in the making. Finally I would like to note that the literature used in the study is limited to those written in English.

[13]The processes of adopting the GDPR was among such challenges

Chapter 2

The Treatment of Biological Materials as 'Data'

1 Data Defined

The terms 'data' and 'information,' though key legal jargons, are often taken for granted and insufficiently, if at all, defined in data protection discourse.[14] Data is habitually used as a synonymous with information. Scholars attribute this dearth in clarity, specifically in laws directly dealing with information concepts, to various factors and contestable assumptions ranging from a simple oversight, to an assumption of obviousness, and to pessimism that the term is incapable of definition, at least a legally workable one.[15]

While it has worked reasonably well in the past the absence of clear definition[16] of the two terms is at its unsustainable stage. The most germane reason for the purpose of this study is the challenge scientific and technological developments[17] introduce to the boundary between information and biological materials—and, in effect, traditional distinction between the message and the medium—which can also trigger application of laws that employ information concepts to biological material.[18]

Outside the legal world, the day to day usages of the two terms seem to draw no clear line of distinction; neither is there a need to make any major differentiation. In their normal

[14]The A29WP as well, in its opinion 4/2007 where it defined the concept of 'personal data', took the term 'data' for grated and had never even asked the question.

[15] Bygrave (2015), p. 107-111

[16] By clear definition it is not meant to necessarily create a distinction between the two terms; clarifying them to be synonyms works as well.

[17] As will be discussed further below, these technological developments include the advancement in ICT and Biotechnology which enabled an ever greater generation of information from biological materials, and making them core constitutive elements of information systems. (Bygrave, 2015, p. 93) In addition, developments in nano-technology and neurology are also blurring the boundaries between technology and human body.

[18] Bygrave (2015), p. 94

parlance, Oxford English Dictionary defines 'data' as 'facts and statistics collected together for reference or analyses'[19] and 'information' as 'facts provided or learned about something or someone.'[20] Even though a first glimpse at these definitions indicates that information is a result of analysis carried out on data, one can also see the usage of the word 'facts' in both definitions which suggests that no serious distinction is aimed to be made. Besides, the thesaurus[21] section of the dictionary puts 'information' and 'data' as synonyms.[22]

In the fields of Informatics and Computer Science, however, a more systematic distinction is drawn between data and information. In these fields, the notion of 'data' usually denotes signs, patterns, characters or symbols which potentially represent something (a process or object) from the 'real world' and, through this representation, may communicate information about that thing.[23]

Expectedly, compared to the vague day to day and legal usage, the distinction made in Informatics is more logical and coherent. The question, however, is would these conceptual wall built in the fields of Informatics and Computer Science be sustainable on the face of the current development in ITC and bio-technology? And even if they continue to work, should the same distinction be made in legislating new or interpreting the existing laws dealing with information concepts? By focusing on data protection law among the latter types of laws, the following sections will strive to address these questions.

[19]Available at: (Link)

[20]Available at: (Link)

[21] The thesaurus also lists other related words like facts, figures, input, documentation and file as synonyms to data/information

[22]Available at: (Link)

[23] Paolo Atzeni *et al*, *Database Systems: Concepts, Languages and Architectures* (McGraw-Hill, 1999) 2;

2 Are Biological Materials Personal Data in the EU Data Protection Regime? (*lex lata*)

2.1 The Existing Legal Regime

2.1.1 The Data Protection Directive

A brief glimpse at the EU Data Protection Directive (DPD) not only fails to answer whether biological materials are considered as personal data but it also makes the answer even fuzzier by its interchanging usage of the words 'data' and 'information.'[24] However, a closer look at the provisions of the DPD indicates absence of intention by its architects to consider biological materials as personal data. Though absence of intention to cover biological materials appears clear, for reasons discussed below, one cannot at the same time, plausibly argue that that was an intentional exclusion either.

First, absence of a clear intention to consider biological materials to be personal data is rooted on how the law and policy in this area generally operates. Professor Bygrave observes:

"[T]he law and policy on data protection have generally tended to operate on the assumption that a distinction exists between data/information on the one hand, and, on the other, the person(s) to which the data/information can be linked."[25]

We see this in the definitions of 'personal data' and/or 'personal information' given in data protection laws.[26] Therefore, paucity of a good indication to treat biological materials as personal data begins from the very definition under Article 2(a) of the DPD. The definition portrays '*humans*' as data subjects to which *information* relates; not humans, or a sample

[24]For instance, recital 26 in the preamble to the DPD uses both 'information' and 'data' in the same context when it tries to delimit the application of data protection principles. This is problematic because, even when human biological materials may be considered as 'data', along the lines of the conceptual distinction between information on one side and data on the other, the directive does not make sense of such distinction.

[25]Bygrave (2010), p. 13

[26]Ibid

taken from them, as information by themselves. It is worth noting, though, that when it tries to further define 'an identifiable person' the directive employs terminologies that relate to the human body. It provides that, in addition to information like a person's identity number, a person can be identified by his physical, physiological or mental identity. Yet, a reference to—say physical identity of a person to identify him, quickly winds up being an information about his physique such as his appearance and not the physical self as such. The same picture can be derived from the preparatory materials towards adoption to the directive.[27] The then EC Commission's commentary[28] to this part of Article 2(a) of the directive, after indicating the typical numerical information[29] as identifying factors, reveals that the definition would also cover data such as appearance, voice, fingerprints and genetic characteristics.[30]

Secondly, other key provisions of the DPD also indicate the absence of a positive intention[31] by the legislature to treat biological materials to be personal data. Some central words and phrases used throughout the directive cannot semantically accommodate human biological material. Words like 'recording' and 'alteration' as set of operations to be performed on personal data under Article 2(b) of the DPD epitomize such inhospitable accommodation. Other instances are under Article 6 whereby personal data is required to be 'accurate' and 'up to date' which presupposes that data could be 'inaccurate' and/or 'out of date', which a biological material cannot be. Similarly, the right to 'rectify' under Articles 10 and 11 presuppose some form of error in recording.

Thirdly, and perhaps more importantly, the crafting of the scope of application of rules of the DPD, under Article 3, cannot comfortably accommodate the application of rules of the directive to human biological material. The directive applies to processing of personal data

[27]Commentary of the Commission, October 1992: COM (92) 422 final—SYN 287, p. 9

[28]Ibid

[29]A person can be identified....indirectly by a telephone number, a car registration number, a social security number, a passport number or by a combination of significant criteria...

[30]Commentary of the Commission, October 1992: COM (92) 422 final—SYN 287, p. 9

[31]By positive intention I mean a deliberate and calculated move from the architects to consider biological materials as personal data.

in two scenarios: (a) wholly or partly by automatic means, and (b) to manual processing of personal data which form/intended to form part of a filing system. At least partly automatic processing of data, which the directive requires under the first scenario, has in mind the use of a device, usually computers, to process information electronically, i.e. when data is computerized. This is precisely what is referred to by the Commission's commentary on this provision.[32]As far as biological materials are concerned one may not, right away, use computers to process blood samples or a swab of specimen of a person. An exposure to a different interpretative framework may be required. The same holds true for the second scenario, i.e., filing system: a file presupposes information recorded on a paper.

2.1.2 The General Data Protection Regulation

Having been invited by the European Council to evaluate the functioning of EU instruments on data protection, as part of the Council's Stockholm Program Notices[33], the EU Commission came up with a proposal for the GDPR in December, 2012.[34] On 12 March 2014, European Parliament made its formal First Reading vote confirming the text of the draft Regulation.[35] EU Justice and Home Affairs ministers reached a general approach on the Regulation at their Council meeting on 15 June, 2015.[36] After months of "trilogue" negotiations, the EU Commission, Parliament and Council of Ministers reached agreement on the GDPR on 15th December, 2015.[37] Following political agreement reached in the "trilogue" the official texts of the Regulation was published in the EU Official Journal on 4 May 2016. While the regulation will enter into force on 24 May 2016, it shall apply from 25 May 2018.[38]

---

[32]Commentary of the Commission, October 1992: COM (92) 422 final—SYN 287, p. 12

[33] The Stockholm Programme — An open and secure Europe serving and protecting citizens, OJ C 115, 4.5.2010, p.1.

[34] COM(2012) 11 final

[35]Bird & Bird, EU Framework Revision: Overview, at: (Link)

[36]Ibid

[37]Ibid

[38] European Commission, Personal Data Protection, available at: (Link)

To examine the position taken by the GDPR on the issue of human biological material, I will analyze, mainly, the official text (of 4 May, 2016). However, to trace the developments on this issue I will also make references to the Commission Proposal (of January 2012), Parliament's first reading (of March, 2014), the Council's general approach (of June, 2015) and the compromise text that resulted from the final trilogue.

The Commission's proposal explicitly mentions the term 'biological samples'[39] in recital 26 of the preamble to the proposed regulation. The mention is made while enumerating the constituents of personal data relating to health. It reads:

> *Personal data relating to health should include... information derived from the testing or examination of a body part or bodily substance, including biological samples...*[40]

Whilst a bold step in separately and explicitly bringing up 'biological samples' which creates a tempting syntax to consider 'biological samples' as personal data relating to health, a closer examination of the recital as a whole shows that it is dealing with information derived from testing or examination of biological samples, not biological samples in themselves. In other words the recital conveys the following meaning: personal data relating to health should not be limited to the information derived from testing/ examination of body part or bodily substance (which require the physical presence of the examinee) but should also include the result of examination of samples when it is taken from examinees the presence of whom is no longer required for examination.

While the same ambiguous syntax is employed in other language versions such as Danish, Swedish and French, Professor Bygrave observes that the German version rules out such ambiguity.[41] In that case, it comes down to a question of interpretation: which language version takes precedence? Recourse to the jurisprudence of the Court of Justice of the EU tells us that the different language versions are all equally authentic and, interpretation of a

[39] The word is mentioned for the first time in EU instruments on data protection.

[40]Recital 26 of the preamble to the Proposed General Data Protection Regulation

[41]Bygrave (2015), p. 6

provision of Community law involves a comparison of the different language versions.[42] The court further notes that every provision of Community law must be placed in its context and interpreted in the light of the provisions of Community law as a whole, regard being had to the objectives thereof and to its state of evolution at the date on which the provision in question is to be applied.[43] Therefore, the task of ascertaining the true meaning of deferring language versions is not simply mechanical, i.e. it does not depend on comparison of the number of versions that avoid the problematic syntax against those which contain such syntax. It should be rooted in the context in which the words are placed, its evolution and the objective of the law as a whole. Seen from this angle it is difficult to claim that the proposed Regulation, indeed, considers biological materials as personal data related to health.

The European Parliament's first reading did not introduce changes to the Commission's proposal in this regard. A small alteration with additional mentions[44] of 'biological samples' came with, first, consolidated text of the Council and the Commission and, latter, with the compromise text. In these versions recital 26 to the preamble of the regulation reads:

> "*Personal data concerning health should include... information derived from the testing or examination of a body part or bodily substance, including genetic data and biological samples....*"[45]

As can be discerned, in this version of the regulation the phrase 'genetic data' is added to the original script. The overall reading of this part of recital 26 would not offer the exact same meaning that the corresponding sentence in the Commission's version did. In that version the phrase 'biological samples' can be meaningfully read back to 'Information

[42] Case-283/81, *CILFIT v Ministry of Health* [1982], Para., 18

[43]Ibid, Para., 20

[44]The compromise text mentions the phrase 'biological samples' at three different instances in the regulation. The first being in recital 26, the other two are made in relation to elaborating and defining 'genetic data' under recital 25(a) and Article 4(10) respectively.

[45]See recital 26 in the preamble to the GDPR (the compromise text)

derived from testing or examination of...' That makes sense because like body parts or body substances, biological samples can also be subjects of the said testing/examination, thus, be carriers of personal information to be derived from them. In addition, referring the phrase 'biological samples' to the 'information derived from testing or examination of…' would be repeating oneself as 'examination of a body part or bodily substance' is already mentioned and biological samples can be considered to be body parts/ bodily substance.

However, the same interpretation wouldn't be logical with addition of 'genetic data' in the later versions of the regulation. That is mainly because genetic data is already a result of analysis of biological materials.[46] Genetic data is generally understood to be information by itself, and while possible it usually is not a subject of testing or examination to derive information, as we do so from body parts/ body substances. Therefore, it creates a temptation to read 'genetic data' and 'biological samples' back to the phrase with which the recital begins: 'personal data concerning health should include...' Otherwise, referring it back to the inner phrase which reads: 'Information derived from testing/examination of...' would end up being, 'information derived from testing/examination of information about heritable characteristics of individuals. That, in turn, ends up being 'Information derived from testing/examination of information.'

While not particularly strong, this can be taken as a reasonable interpretation of the wordings of the compromise text. But it still remains ambiguous at this point. This interpretation also advances the attainment of the general objectives[47] of the regulation set out by the Commission, particularly the first objective: helping citizens to be in control of their data.[48] After all, the very conception of privacy is ingrained in the protection of personal integrity which, at some level, requires extending protection to our biological materials.

[46]Recital 25(a) and Article 4(10) of the compromise text of the regulation clearly testify to the fact that genetic data results from the analysis of biological samples.

[47]The Commission sets out three general objectives for the regulation, See The Proposal for GDPR, P. 102

[48] Some commentators, though, have argued these objectives are based on fallacious assumptions, thus, unattainable. See, Koops(2014), '*The trouble with European data protection*'

However, towards the end of writing this study, the official text of the Regulation is published in EU Official Journal on 4 May 2016.[49] Recital 35 in the preamble to the official text of the regulation clarifies some of the issues raised with in recital 26 of the previous versions. The relevant part of the recital reads:

> "Personal data concerning health should include … information derived from the testing or examination of a body part or bodily substance, including *from* genetic data and biological samples[50]*... "* (emphasis added)

The addition of the preposition '*from*' now makes it difficult to read 'biological samples' back to the beginning of the recital. It should be read with the phrase 'information derived from testing or examination of...' This implies the absence of positive intention by the architects of the regulation to consider biological samples to be personal data. The previous version can, therefore, be considered as a result of poor draftsman-ship.

Having said this much about the DPD and the GDPR I will now briefly turn to the status of biological materials under European case laws, and national legislations. The focus of the study being on the legal regime at European level the coverage on national legislation will only be brief. As far as national laws are concerned they appear to be divided along geographic lines. Many western European countries tend to adopt the view that biological materials are not personal data while some eastern European countries have taken the opposite stance. Bulgaria, Estonia, Latvia and Romania are among eastern European countries that recognize body samples as data in contrast with other western European countries like Spain, Portugal and Germany.[51] Outside Europe the Australian state of South

[49] 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)

[50] Recital 35 in the preamble to the GDPR (EU Council's Position with the view of adoption, 6 April, 2016)

[51]See, Bygrave (2010), p. 16-17 for references.

New Wales's privacy and information legislations clearly include bodily samples in their definition of personal information.[52]

As was the case for data protection in general, case law on the issue of 'biological materials as data' has not been abundant. While there is considerable number of case law relating to data protection today many of them have hardly shed any light on the issue of bio-samples as data. That could be attributed, at least partly, to the level of awareness of the European population regarding bio-banks in general; not just what bio-banks are used for or how they may affect fundamental rights, but the very fact that they exist. One study of the European Commission found that more than two third (67%) of Europeans have never heard about the term itself.[53] Only 2% of the population have actively inquired in to and searched about bio-banks.[54] As awareness rises on what bio-banks are, how they are used, and their effects on privacy, it can be expected to lead to privacy litigations which would involve biological materials.

Among the few instances in which courts dealt with this issue are in the cases of *S and Marper v United Kingdom*[55] handed down by the European Court of Human Rights and, the decision of Norwegian Data Inspectorate.

In *Marper* the European Court of Human Rights, essentially, ruled that retention of fingerprints, cellular samples and DNA profiles of individuals arrested but who are later acquitted or have charges against them dropped is a disproportionate interference to their right to privacy under Article 8 of the European Convention on Human Rights. That being the chief finding of the court in this judgment, the court has also directly, though scarcely,

---

[52] section 4(2) of Privacy and Personal Information Protection Act 1998, section 5(2) of the Health Records and Information Privacy Act 2002 and the Government Information (Open Access) Act 2009, Schedule 4, clause 4(2)

[53] EU Commission(2012), *Bio-banks for Europe: A challenge for governance,* P. 24

[54] Ibid, p. 25

[55] *S and Marper v United Kingdom,* European Court of Human Rights,(App no 30562/04 and 30566/04), 4 December 2008

addressed the issue of human tissue samples. It found that cellular samples constitute personal data within the meaning of Data Protection Convention:

> *The Court notes at the outset that all three categories of the personal information retained by the authorities in the present cases, namely fingerprints, 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 [UK] 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*.[56]

While a remarkable judicial activism, the effect of this view in the judgment is limited in a number of ways. It figured only marginally because the court did not need to delve in to the issue of biological material as application of Article 8 ECHR, on which the judgment is based, does not turn upon whether 'data' or 'information' are/is processed but on whether there is interference with the right to respect for privacy. The court does not also have a legal mandate of interpreting the Data Protection Convention.[57]

It is also worth mentioning here that in prior litigation of the case in the UK by the House of Lords the issue of bio-samples as data is directly touched up on by Baroness Hale. She argued that the same privacy principles should apply to all the three (fingerprints, DNA profiles and cellular samples), essentially, because they are all kept for and as 'information.' Those are her words:

> "But the only reason that they [samples] are taken or kept is for the information which they contain. They are not kept for their intrinsic value as mouth swabs, hairs or whatever. They are kept because they contain the individual's unique genetic code within them. They are kept as information about that person and nothing else. Fingerprints and profiles are undoubtedly information. The same privacy principles should apply to all three.[58]"

[56] *S and Marper Vs UK*, para. 68

[57]For detailed analysis of this decision, see Bygrave (2010) p. 7-13

[58] *S, Regina (on application of) v South Yorkshire Police,* [2004], Para.70

As will be discussed in the next section, Hale's point forms one of the basic arguments put forth in favor of considering bio-samples to be data/information.

3 Should Biological Materials be treated as Personal Data (*lex ferenda*)?

There is no consensus on the issue of whether human biological materials should be treated as personal data. Some scholars, commentators and agencies enforcing data protection laws have taken the view that personal data should not be seen to include biological materials for the purposes of data protection laws. The Article 29 Working Party[59] and the UK's Information Commissioner's Office (ICO)[60] are cases in point. In its opinion where it clarifies the concept of personal data under the DPD, the Working party makes a clear distinction between biometric data—which it rightly considers as personal data—and, human tissue samples from which biometric data is extracted, which it is opined not to constitute personal data. In the Working Party's words:

> *Human tissue samples (like a blood sample) are themselves sources out of which biometric data are extracted, but they are not biometric data themselves (as for instance a pattern for fingerprints is biometric data, but the finger itself is not). Therefore the extraction of information from the samples is collection of personal data, to which the rules of the Directive apply.*[61]

In a similar way, the official view from the UK's Information Commissioner is reported to be analogous: a sample is not treated as personal data, 'because it is physical material'.[62]

On the other hand, even though much of the data protection law and policy have been operating on such distinction scholars[63] have questioned the logic underlying the distinction

[59] The Article 29 Working party (A29WP) is an independent advisory body established by the Article 29 of the EU Data Protection Directive

[60] The ICO is the UK's independent body set up to uphold information rights in general, including those under the UK Data Protection Act.

[61] A29WP, Opinion 4/2007, p. 9

[62] Ashgate(2004), in Deryck Beyleveld *et al* eds. P. 428

between human biological materials on the one hand and personal data on the other. Those pushing the view that biological material may be personal data or information tend to pay more regard to pragmatic considerations such as the need to fill lacunae in bio-bank regulation, the growing ease with which persons can be identified from biological material, and the fact that such material is often only stored for generating information.[64] Others who take the view that a biological material does not constitute personal data depend on conceptual logic claiming that "data is a formalized representation of objects or processes, while information comprises a cognitive element involving comprehension of the representation."[65] In the following sections I will analyze whether such conceptual distinction still makes sense, at least as far as (human) biological materials are concerned, in relation to recent developments in the field of bio-technology.

3.1 The Conceptual Framework: Does it still make Sense?

3.1.1 DNA: the Game Changer

The discovery of the structure and basic nature of DNA (deoxyribonucleic acid) as carrier of human genetic information around mid-20th century brought about significant development on how we understand the operation of life forms. It has been argued that the discovery of DNA, and our understanding of its structure and functioning may well be the most important discovery of the last century.[66] The effect of the discovery on scientific and medical progress has been enormous, whether it involves the identification of our genes that trigger major diseases or the creation and manufacture of drugs to treat these diseases.[67]

---

[63] Bygrave (2010); Bygrave (2014); Taylor (2012), Chapter 7; Ploeg (2007)

[64] Bygrave (2015) p. 7, Bygrave (2010) p. 8-9

[65] Ibid, p. 6-7

[66] Murnaghan (2016), available at Explore DNA, Available at: (Link)

[67] Ibid

Among the other noteworthy effects of this discovery (reinforced later by the genome project[68]) is the characterization of DNA as a recipe of life; a carrier of information based on which our cells make the necessary protein in our body. That means the very essence of all living cells which make up a human person are the products of those information. But before that analysis it is important to say few words on the meaning and nature of the DNA to put the discussion in context.

Our bodies are made out of billions of individual cells, and DNA is the control center of each and every cell.[69] DNA is the hereditary material in humans and almost all other organisms. Nearly every cell in a person's body has the same DNA.[70] Therefore, almost every cell in our body houses complete set of our hereditary materials, i.e., the genome.

On a deeper level, DNA consists of a strand of four nucleotides called adenine, guanine, cytosine, and thymine, commonly abbreviated to A, G, C, and T respectively.[71] A particular arrangement of these nucleotides forms up a gene. Genes specify the kinds of proteins that are made by cells.[72] That means, the sequence of the nucleotides are read to make a particular type of protein that our body needs. It is from that information that proteins are made.

Almost everything in the body, from hair to hormones, is either made of proteins or made by them.[73] Therefore, as protein forms the building blocks of our body it literally means that we are made up of information read from our DNA, the arrangement of nucleotides.

---

[68] The Human Genome Project (HGP), undertaken from 1990 - 2003 with billions of dollars involving multiple continents, was an international scientific research project with the goal of determining the sequence of chemical base pairs which make up human DNA, and of identifying and mapping all of the genes of the human genome from both a physical and functional standpoint.

[69] Calladine, et el (2004) p.3

[70] Some cells, like the red blood cell, do not have nucleus, thus, DNA (Ridley, 1999, P.6)

[71] Amos(2005), p. 6

[72] Berg JM (2002), Chapter 5

[73] Ridley, (1999) P.7

That is why Matt Ridley wrote "the idea of the genome as a book is not, strictly speaking, even a metaphor. It is literally true."[74]

This striking scientific discovery about our body is at odds with the traditional conception of distinguishing data as (medium representing reality) as opposed to information (comprehension of the representation), at least as far as the body is concerned. The human body itself is a construct of information; information which instructed the formation of proteins from which body is formed.[75] The conceptual rigor, therefore, begins to crumble when we closely scrutinize the human DNA.

In fact, DNA as a carrier of information is not limited to carrying genetic information for the formation of our body; a scientific breakthrough has made it possible to carry external large size digital information for a long time.[76] But that development still remains nascent.

3.1.2 Other Developments in Biotechnology and Beyond

In addition to the scientific facts revealed about our DNA, the conceptual distinction between data and information is also challenged by multiple other developments that blur a clear boundary between biology and technology.

First, after the Human Genome Project, another initiative labelled 'America's next big thing'[77]in neuroscience research, called the 'BRAIN' (Brain Research through Advancing Innovative Neuroethologies) has been announced by President Obama in his State of the

[74] Ridley (1999) p.6

[75]It may be important to note here that my argument is only limited to biological materials. The conceptual distinction, otherwise, still makes full sense elsewhere.

[76]See, Independent, *Single DNA molecule could store information for a million years following scientific breakthrough*, 17th August, 2015.

[77]Such project, though, is not of interest only in the United States of America; the European Commission has almost simultaneously announced the Human Brain Project with an award of 1.19 billion Euros. (See, Kaku (2014), p. 250)

Union address of January 2013.[78] The BRAIN initiative aims to decode the tens of thousands of connections between each of the ~86 billion neurons[79] that form the basis of human brain.[80] That means, as the Human Genome Project sequenced all our genes, the BRAIN initiative will map all of our neurons. That can be said to be the general goal of the initiative.

The unstated goal of this initiative, the part directly germane to this study, is eloquently described by Dr. Michio Kaku, Professor of Theoretical Physics at City University of New York in his 2014 book titled 'The Future of the Mind.'[81] The expected main output of this project is what scientists call a *connectome*: a comprehensive map of neural connections in the brain which encodes all our memories, dreams, hopes and desires, perhaps, on a CD. This raises very important questions: by putting together a CD of a person's connectome with his genome, are scientists creating, in some sense, immortality?[82] Because even after a person has passed away his body could be revived from his genome; while his consciousness from his connectome. That means, we can continue to live even after we are dead: as information. This possibility that we can still continue to live as information tempts us to conclude that we are nothing but information.

Secondly, the undergoing various forms of 'human enhancement projects[83]' are clouding the boundary between human body and technology. Our body may no longer be limited to

[78] See**,** Isabelle Abbey, News and Views: The Brain Activity Mapping Project – What's the plan? April 24, 2013**.** Available at: (Link)

[79]Neurons are nerve cells that carry information between the brain and other parts of the body (Cambridge Dictionaries Online)

[80]Ibid

[81]Kaku (2014), p. 252

[82]Ibid.

[83]In the context of engineering, human enhancement can be defined as the application of technology to overcome physical or mental limitations of the body, resulting in the temporary or permanent augmentation of a person's abilities and features (See, *Human Enhancement,* Dartmouth Journal of Undergraduate Science, In Fall 2013)

what it is today.[84] Its shape, composition and, as a result, its capabilities are radically changing. It is now clear that "human enhancement" is a reality and not just a product of science fiction.[85] Even more so, as technological advances will imminently provide various devices that will interface with the human body in various ways.[86]

Thirdly, the steadily growing accumulation of human biological samples in bio-banks[87], and the increased deployment of biometric technologies in every sector are 'infromationalizing' the human body by converting features of it in to processable digital data.The upsurge in coverage, sophistication and use of bio banks, is spurred to a large extent due to the advances in genetic science.[88] The need for identification/verification of persons in both public (such as in forensic investigations) and private (in cases like private security) is largely the reason for the expansion in deployment of biometric technologies. Regardless of the reasons for their upsurge they have a clear common effect: conversion of particular aspects of physical existence into electronic data and digitally processable information.

All of these developments: from the sequencing of our genome, to the future mapping of our neurons, to the various human enhancement initiatives, and to our continued existence in the form of biometric information undoubtedly challenge the conceptual separation between the human body on the one hand, and information about it on the other.

Dr. Irma Ploeg, convincingly, suggests that this should be seen as something more profound than constituting yet one more instance of the collection of "personal

[84]As a naturally (biologically) constituted being with natural organs, muscles, bones and bodily fluids.

[85]The Guardian: *Yes, nano science can enhance humans – but ethical guidelines must be agreed,* Monday 3 June 2013

[86]Ibid; an article in Science Magazine exemplified how machines can interact with living brains to allow wireless changes in behavior by the implantation of devices directly into the brains of mice. These devices could then be remotely controlled to activate different parts of the brain using light. (*Science Magazine, Injectable, Cellular-Scale Optoelectronics with Applications for Wireless Optogenetics,*12 Apr 2013 (www.science.sciencemag.org))

[87] Bio banks may exist in any forms; be it, tissue, blood, cell material, skin, gamete, or embryo banks.

[88] Bygrave (2010), p.3

information", as is more commonly done. Rather, the human body is implicated in a process of co-evolution with technology, information technologies in particular.[89] It calls for a different conceptualization of bodily existence: body as information.[90]

3.2 Pragmatic and Other Considerations

In the previous section it is argued that the conceptual distinction between biological material and information can no longer be logically defended for all the reasons discussed therein. In this section, I will turn to the more pragmatic and persuasive reasons for extending the definition of 'personal data' to find a room for biological materials.

3.2.1 Indistinguishable Interpretive Potential[91]

If one is concerned about practically preventing adverse effects on the right to privacy, what matters most is the interpretive potential of data/source i.e. the ability to generate information that can be linked, not just the assumed availability of identifiable information. If any concerning, from privacy view point, identifiable information can easily and readily be generated from a given source —which more often is the case for biological samples— then that raises as much privacy concerns as the information derived from them would. We can consider two important, but related reasons to substantiate this sameness in interpretation potential between the two.

First, if interpretation[92] is the reason for the distinction, even recorded information will undergo an interpretation before it informs. Taylor observes that: it remains the case that data (as recorded information) must always be interpreted before its meaning can be understood: records must be read. If the privacy protection established by the Directive extends to include the physical record of information, then the viability of any division

[89] Irma (2007), p. 47

[90] Over the past century developments in the medical Sciences have resulted in various body ontologies like 'the endocrinological body' (in the early twentieth century) whereby the body is viewed as just biochemical entity. (Irma 2002)

[91] By 'interpretive potential' I am referring to the ability to generate (potentially) identifiable information.

[92] By 'interpretation' I mean mechanisms and processes that may be employed to derive information from biological materials.

between (biological) sample and information built upon the former's need for subsequent interpretation crumbles.[93]

Secondly, even if recorded information might be said to have an imminent and easy potential to inform than a biological material before it is interpreted, this would not lead to the conclusion that the relative ease in accessibility of recorded information puts right to privacy any more vulnerable than biological materials. It all depends on the availability of the necessary interpretive framework to derive readily accessible information from the samples. A western person, born and raised in the west, may not be able to be informed by having access to 'information' written in an eastern script—say Mandarin. But that does not, in any way, mean that the 'Mandarin text' is not recorded information. It just means that, for that text to inform the necessary framework should be in place: the skills to read and understand Mandarin.

Thus, recorded information and biological samples have an indistinguishable potential of putting right to privacy in jeopardy. In some situations, however, a concern from biological samples could be much worse. Interpreted information may be manipulated, if necessary, to meet certain privacy standards while biological materials will always be available to give away any information in the open. While the manipulation of data may seek to make certain information more accessible, it might also seek to obscure it (e.g. through coding), and the source data may remain interpretable in any event.[94] In this regard, Taylor argues that even information, not just samples, can be subjected to new interpretation, thus, sharp distinction should not be drawn between recorded information and bio-samples.[95]

While Taylor's argument is valid it should be noted, however, that bio samples are more susceptible for new form of interpretation as they are often kept for interpretation and only for interpretation. That makes, in some situations, biological materials even more worrisome in terms of privacy than information derived from analysis of such materials.

[93] Taylor (2012)p. 162

[94] Ibid, p. 163

[95] Ibid, p. 164

Similarly, the interchangeable usage of the words 'information' and 'data' both in the law and policy circles— including in the DPD— and in our day to day usage is yet another tribute to similar effects that they produce implying absence of a real reason to distinguish the two. Two reasons are worth mentioning for such interchangeability. The first one explains why we, hitherto, use the two words interchangeably, and the second pertains to why we will, perhaps, continue to do so even more in the future.

First, information derived from interpretation of data can then be recast and used as data for another interpretation in a way that we are tempted to use the two words interchangeability.[96] From a given national census, for instance, sex and age 'data' can be used to derive 'information' about the percentage of the youth in a relevant population which can, in turn, be used as 'data' for youth centered policy making. In the same token, information derived from biological materials can be used for another analysis as data.

Secondly, pervasive, repeated and systematic extraction of information from human biological materials would eventually end up making the bio-samples themselves 'information' mainly because the extraction is of such extensive nature and the sole reason they are stored is for information. This trend can be paralleled with the gradual change in meaning of the search engine 'Google'. Because of large scale usage of this service 'searching' on the web by authoring some key words came to be analogous as 'Googling.' This development came from the repeated and extensive use of 'Google' for indexation even if Google still remains just one search engine provider and the term does not have any semantics indicating 'search.' In a similar way, continuous and pervasive derivation of information from biological materials means that, more and more tempting interchangeable usage of the two words. So, a time may come when we will call 'bio-sample' as information not just data. It all depends on how easily accessible the interpretative frameworks are and how frequently we use them.

[96] Taylor (2012), p. 42

3.2.2 Enhancing Bio-bank Regulation

The other major benefit expected from the inclusion of biological materials in to the concept of personal data is the anticipation to fill the regulatory vacuum in bio-banks. What makes this regulatory vacuum all the more germane to data protection discourse is the fact that it is manifested in the incapacity to effectively preserves fundamental rights of privacy and data protection of participants even though such is one of the primary objectives of bio-bank regulations. In this regard, one of the EU Commission's study on Bio-bank governance notes 'one of the main challenges has been, and still is, to identify ways to protect the autonomy and dignity of patients and research participants and their fundamental rights (e.g. private life and data protection, especially in case of loss of control on personal data/data misuse, discrimination) with fostering the public interest in carrying out medical research to address the central public health challenges (such as cancer, cardiovascular and metabolic diseases.)'[97] The same study reiterates absence of clear legal framework governing bio-banks as one of the major problems for the imbalance against protection of fundamental rights.[98] With relatively comprehensive rules and well established enforcement mechanism data protection laws can serve as a better mechanism even though the latter also have their own limitations.[99]

3.2.3 Just 'About Us' or but not 'Us' (A Moral Plea)

As it stands today the existing data protection regime in the EU protects information that *relates to* us but not, strictly speaking, us. Even by a layman standard of what is right and wrong, leaving out bio-samples will be wrong. To make full sense of how morally questionable the current system is, one needs only to consider two facts against which this moral claim should be assessed. One is the fact that the starting point of discussions on the right to privacy has usually been a concern for bodily integrity. The division between informational privacy and bodily privacy are made fictitious by technological development,

[97] EU Commission(2012), *Bio-banks for Europe: A challenge for governance,* P. 45

[98]Ibid, p.46-48

[99]See, Bygrave (2010), p. 21-22, for details and references on similar problems of some European national bio-banks regulations

especially since the past decade. In this regard, the Australian Office of Federal Privacy Commissioner, back in 2002, rightly noted:

> *... an attempt to maintain a clear demarcation between different types of privacy protection may be problematic in light of new technologies which involve the merging of biology, mathematics and computer science, namely, biometrics and bioinformatics. Such developments give rise to new forms of body templates or records which further blur the distinction between personal information and its source in individual humans, rendering the concepts of information privacy and bodily privacy inherently interrelated*.[100]

Secondly, on the face of such division the regulatory landscape pertaining to bio-banks has largely been uncoordinated and ineffective as noted above. Therefore, not only does this fact stand in contrast with the original conception of privacy thus failing the very essence of its inception, the human body is also failed by disarrays in regulation of bio-banks.

Against these two backgrounds alone, is it morally indefensible to protect information about individuals but not individuals themselves, or a sample taken from them. The human body or a sample taken from it is one of the most sacred representations of one self. To argue that a fingerprint represents the finger while a sample doesn't represent the person is not only morally questionable but also logically weak. Distinction should also be made between the human body/sample as source of data/medium and other sources of data as integrity and privacy is often an issue when human body is involved.

4 The Consequences of Treating Biological Materials as Personal Data

Despite crumbling conceptual rigor that distinguishes human biological materials from data/information, and various pragmatic considerations that increasingly challenge such

[100] ALRC and AHEC, (2003), *Essentially Yours*, p.280

distinction, collapsing differences that has been maintained in the regulatory discourse for such a long time is not without its own drawbacks.

4.1 Over Stretching the Scope of Data Protection Laws

The inclusion of a new subject matter in to the scope of application of data protection law, to the least, demands a closer look at the existing subjects of the law to see whether it properly fits with the law's regulatory apparatus. Data protection law already suffers from regulatory overreaching in the sense that its rules tend to apply *prima facie* to a wide range of activities with relatively scant chance of being respected, let alone enforced.[101] The Data Protection Directive is, for instance, said to have a long arm with application to multiple actors based outside the European Union.[102]

Article 4(1) (c) of the Data Protection Directive epitomizes one such long arm. This provision subjects any controller located anywhere in the world to European data protection regime when it utilizes an equipment situated in any member state for the purpose of processing personal data.[103] The General Data Protection Regulation, perhaps, does more than the directive in this regard.[104]

4.2 Centrality of Consent

The other problem in the inclusion of biological materials in to the scope of data protection regime comes from the inadequacy of the current rules to meet the normative position of consent in the laws currently concerned with regulation of biological materials. The fundamental principle that underpins the governance framework of human biological

[101] Bygrave(2010), p. 22

[102] See, Lokke Moerel (2011)

[103] Bygrave(2014), p. 202

[104] The Regulation applies to controllers not established in the Union when they process personal data of European residents in relation to the offering of goods and services to them and monitoring of their behaviour (Article 3(2)). The Parliament's version of the regulation, which has also made to the compromise text, even goes on saying that the goods and services need not be offered for consideration (The Parliament's reading and the Compromise text of the GDPR, Article 3(2)). The final version of the regulation did not change the compromise text.

materials in general is the need to obtain voluntary and informed consent of participants. The history of how biological materials were governed—such as by the European Convention on Human Rights and Biomedicine, and Declaration of Helsinki[105] show that consent is unequivocally important as it occupies a central normative position. The Convention on Human Rights and Biomedicine stipulates that an intervention in the health field may *only* be carried out after the person concerned has given *free and informed consent* to it. This person shall beforehand be given appropriate information as to the purpose and nature of the intervention as well as on its consequences and risks.[106] The interests and welfare of the human being shall prevail over the sole interest of society or science.[107] In addition to securing free and informed consent for the purposes of medical research the convention requires other safe guards like making sure that there is no alternative of comparable effectiveness to research on humans.[108]

In this regard, the Data Protection Directive or the Regulation are too liberal to accommodate what is customarily and legally expected if biological materials were to be governed by these regimes. That requires the role of consent under the directive and the General Data Protection Regulation to be seen more closely.

4.2.1 Does Consent Play Central Role under the Current EU Data Protection Regime?

Broadly speaking, data subject's consent is one of many control mechanisms[109] in which data subjects, as active actors in data protection laws[110], influence the data processing operations of controllers. Though there are some non-negligible reasons, in particular for

[105] World Medical Association (WMA), World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects, 2008.

[106] Council of Europe, Convention for the Protection of Human Rights and Biomedicine, Article 5

[107] Ibid, Article 2

[108] Ibid, Article 16

[109] Other control mechanisms in which data subjects can influence processing of personal data can be: opposing a particular processing or withdrawing consent.

[110] We have two additional main actors in the operative sphere: DPAs and controllers (Bygrave 2014, p. 18-19)

sensitive personal data, more convincing evidences suggest that consent does not play any central role in the existing data protection regime. There are, however, more stringent requirements for consent of the data subject with regard to processing sensitive data. In principle, processing sensitive personal data is prohibited. In addition, the jurisprudence of the European Court of Human Rights (ECtHR) in some of the cases—such as *Z v Finland* and *MS v Sweden*—suggest normative importance of data subjects' consent regarding sensitive data, particularly, medical information. Thus the problem can, somehow, be mitigated by the fact that consent enjoys relative central role under the directive when with regard to sensitive data. That is because biological materials would most probably belong to the category of sensitive data as data concerning health under article 8(1) of the directive.

Generally, however, under articles 7 & 8 of the DPD, consent is not only just one precondition among the alternatives for legitimate processing, member states are also allowed to introduce new grounds for reasons of substantial public interest.[111] Similarly, the EU Charter of Fundamental Rights provides: personal data can be processed "*on the basis of the consent of the person concerned or some other legitimate basis laid down by law*."[112] While consent is expressly mentioned, the Charter makes it clear that personal data can be processed on the basis of other legitimate grounds laid by law.

In addition, from a pragmatic view point the DPD incentivizes data controllers to first utilize other preconditions—such as the one under article 7(f)—and employ consent when a processing exercise can't be justified under those grounds. This flows from the cost and delay involved from securing consent and, the desire to avoid the possibility of refusal by the data subject.

Though all these facts demonstrate absence of normative priority, a closer look at, at least some of the preconditions, tells us that they are framed on the assumption that 'if the data subjects were asked to consent, they would have agreed to the processing.' The

[111] DPD, Article 8(4)

[112] EU Charter of Fundamental Rights, Article 8(2)

preconditions like 'necessary to protect vital interests of the data subject' and 'necessary for performance of contract in which the data subject is a party' are examples in point. Therefore, I would argue, that the other preconditions also aren't completely devoid of an element of consent. Consent can still be read in to them in its broadest and indirect/implied sense.

However, what is problematic is not just that consent does not play a central role under the existing regime; there are also convincing arguments against a central role of consent as a precondition for data processing. First, there are legal problems in properly delineating the requirements of consent, for instance, how informed should consent be, for instance, under article 2(h) of the DPD. Secondly, the degree of choice presupposed by consent mechanisms will often not be present for certain services or products, particularly those offered by data controllers in a monopoly (or near-monopoly) position.[113] Thirdly, despite the requirements of informed consent and notification (for instance articles 10&11 of DPD) controllers will typically have greater knowledge about their data processing operations than will the subjects.[114] The asymmetry will further weaken the 'informed' nature of data subject's consent. Finally, problems of consensual exhaustion, laxity and apathy – in addition to ignorance and myopia – can reduce the amount of care that data subjects invest in their decisions of whether or not to consent.[115]

Therefore, not only is it doubtful that consent plays a central role in the processing of personal data—including sensitive data—it is also, arguably, not desirable that it plays such a central role. Yet, it remains central in other laws traditionally concerned with human biological materials. Thus, the extension of the DPD or the GDPR[116] to biological materials only poorly meets the central normative position of 'consent' in laws currently governing biological materials. As indicated earlier, this problem can, somehow, be mitigated by the

[113] Bygrave and Schartum (2009), p.160

[114] Ibid. p.160-161

[115] Ibid. p.161

[116] With some clarifications on the requirement of 'consent' the Regulation remains structurally the same with regard to the normative position of consent as a ground of processing personal data.

fact that consent enjoys relative central role under the directive when it comes to sensitive data, the category to which biological materials would most probably belong.

4.3 Enforcement

Yet another major concern in trying to extend the scope of data protection regime is the fear that the enforcement of the law, that includes biological materials, would require strong Data Protection Authorities (DPAs) with additional competence to handle the peculiarities of biological materials. This problem becomes even more alarming because the ability of data protection authorities to ensure effective compliance of the law is already under pressure as they are chronically under-resourced.[117] The addition of biological materials in their task sheet thus fuels the difficulty. DPAs will not only the need additional material resources, but also personnel with broad and interdisciplinary professional background.

[117] Bygrave (2010), p. 22

Chapter 3

The Criterion of Identifiability

In Chapter One, we identified the two words *'information'* capable of *'identifying'* the data subject to be the building blocks in defining personal data under the current EU data protection regime. We then analyzed the first of these building blocks in the subsequent chapter. This chapter turns to the analysis of major problems associated with the usage and application of the criterion of '*identifiability'* as the other essential constituent of the definition.[118]

1 Identifiability: Briefly Defined

The criterion of 'identifiability' forms the second essential ingredient in defining personal data in the current and *en route* European data protection rules.[119] In this definition 'identifiability' serves as a qualifying factor: it qualifies 'information/data' which is being processed as having certain capacity. It posits that *Information/data* should assume a particular capacity to be personal data, thus, the subject of data protection law. It should be able to '*identify*' the data subject, even if identification happens only with assistance from other piece of information/data. The criterion of 'identifiability' serves, therefore, as differentiator. It, in essence, separates data/information that is subject to data protection law from those that fall outside its scope.

[118] It is worth clarifying, here, that the mainstream literature on this area presents the analysis of the definition of personal data as containing more elements than the two under analysis in this research: '*information'* capable of '*identifying'* the data subject. In that literature, including in A29WP's opinion analyzing the definition of personal data, we see the criteria of '*relating to'* and *'physical person'* analyzed. I decided to leave the latter criterion simply because it doesn't involve any significant critical discussion once we clarify that the law only applies to natural persons as opposed to legal persons. Likewise, I did not discuss the criterion of data *'relating to'* the data subject as this requirement can, in most cases be subsumed in the criterion of *'identifiablity'* in a sense that data that is capable of identifying a person can be assumed to also relate to the data subject. (See, Bygrave (2014), p.130-131)

[119] The definition, structurally, remains the same under the General Data Protection Regulation. (See, Article 4(1) therein)

Identifiability denotes a state of being identified (actually) or identifiable (potentially) by a given data/information. The term 'Identified' 'requires elements which describe a person in such a way that he or she is distinguishable from all other persons and recognizable as an individual.'[120] 'Identifiable', on the other hand, entails piece of information [that] contains elements of identification through which the person can be identified, directly or indirectly.'[121]Thus, essentially, identifiability signifies the ability to distinguish a person directly based on pre-collected information or indirectly by pairing the information at hand with other auxiliary data/information. Distinguishing a person is normally achieved through pieces of information that hold particularly privileged and close relationship with a particular individual.[122] These can be a name, identification number, location data, online identifiers of a person or other indirect identifiers like a person's physical, physiological, mental, genetic, economic, or social identity.[123]

The then EC Commission's commentary on article 2(a) of the DPD reveals the legislator's intent to refer to identifiers which are, more or less, 'nominative data/information.'[124] However, as broached above, the General Data Protection Regulation gives recognition to other identifiers such as 'online identifiers.'[125]

The Data Protection Directive, though, doesn't apply to *any* potentially identifiable information. Account should be taken of all the means *likely reasonably* to be used by the

[120] European Union Agency for Fundamental Rights (2014), p. 39

[121] Ibid, p. 40

[122]A29WP, Opinion 4/2007, p. 12

[123] Ibid, p. 12-15; also see, Article 2(a) of DPD & Article 4(1) of the proposed GDPR (Compromise text)

[124] The commission's commentary reads: "a person may be identified directly by name or indirectly by a telephone number, a car registration number, a social security number, a passport number or by a combination of significant criteria which allows him to be recognized by narrowing down the group to which he belongs (age, occupation, place of residence, etc.). The definition would also cover data such as appearance, voice, fingerprints or genetic characteristics." These identifiers are nominative in a sense that they, readily or after some effort, lead to the name of the data subject. See also, Costa & Poullet (2012), p. 255

[125] The GDPR (Compromise text), Article 4(1)

controller or any other person, in order to identify the data subject.[126] These dual criteria further reduce data/information from being subject of the directive by excluding cases of only theoretical identification. The term '*likely*' can be understood to suggest *probability* of identification while the term *'reasonably'* points to the *difficulty* (in terms of legality, time and other resources) involved in identification.[127] Thus, the intent behind identification, the way the processing is structured, the advantage expected by the controller, the interests at stake for the individuals, as well as the risk of organizational dysfunctions, the state of the art in technology at the time of the processing and the possibilities for development during the period for which the data will be processed, the duration of processing and other relevant factors should be considered before a data is said to be 'identifiable.'[128] However, it does not matter, for the purposes of the Directive, who is able to link a person to a data: the data controller or *any* other persons are legally relevant agents of identification.[129]

2 Major Problems Associated with the Usage and Application of the criterion of Identifiability

The analysis of major problems with the criterion of 'identifiability' are made under two sub-sections. In the first subsection I will examine the problems in which individuals are targeted through their data even if they are not/ need not be identified. I call this: pre-identification problems. In the second sub-section investigation of problems with the application of the criterion will be made by assuming 'identifiability' is a fairly proper mechanism to protect privacy and related interests. I call this: post-identification problems.

[126] See, recital 26 in the preamble to the Data Protection Directive.

[127] Bygrave (2002), p. 44

[128] A29WP, Opinion 4/2007, p. 15-17; See also, recital 23 in the preamble to the draft GDPR (Compromise text)

[129] See, Bygrave (2002), p. 45. But this part of recital 26 is not faithfully transposed in some country's national laws like the UK. For the purpose of the latter, only the controller is a relevant agent of identification. (Article 1 section 1 of the UK Data Protection Act) Nevertheless, the relevance of such differing transposition will no longer be relevant once the General Data Protection Regulation becomes operational.

The importance of the second analysis is further strengthened with the fact that the GDPR also operates with structurally the same definition of personal data.

2.1 Pre Identification Interests

2.1.1 Data need not 'Identify' to Cause Harm

The burgeoning importance of personal data in today's internet economy has lead to its characterization in different decisive ways. It has been mainly called 'the new oil'[130], thus, 'the next big thing[131]' and 'representative of a post-industrial opportunity.'[132]

These are all because personal data is generating a new wave of opportunity for economic and societal value creation.[133] Mining and analyzing personal data gives us the ability to understand and even predict where humans focus their attention and activity at the individual, group and global level.[134] Once mined and analyzed, firms use data to support individualized service delivery business models that can be monetized; governments employ it to provide critical public services more efficiently and effectively; researchers accelerate the development of new drugs and treatment protocols, and end users benefit from free, personalized consumer experiences such as Internet search, social networking or buying recommendations.[135]

These benefits, however, are not always in harmony with other societal values we uphold. They are often realized at the expense of our central values and fundamental rights as human beings: privacy, integrity, autonomy and related values. This is where data protection laws come to the picture: they strive to create a mechanism in which we reap the

[130] Meglena Kuneva (European Consumer Commissioner), *Keynote Speech* (SPEECH/09/156), 31 March 2009

[131] Perry Rotella, *Is Data The New Oil?,* Forbes, April 2, 2012

[132] World Economic Forum (2011), Personal Data: the Emergence of New Asset Class, p. 5

[133] Ibid

[134] Ibid

[135] Ibid

economic and related social benefits of modern data processing without unduly compromising our fundamental rights and values. Nonetheless, data protection law applies only as far as the data under consideration relates to an identifiable individual. And when identifiability is strictly defined[136] such as — 'as referring to nominative data'—much of 'personalized' data that relates to individuals remains outside the scope of protection offered by data protection laws. That means, while the regime of data protection law is there to mitigate some of the risks posed by modern data processing practices to our fundamental rights and core values, some of the data processed under processing are not necessarily 'personal data' even if they continue to adversely affect fundamental rights intended to be upheld by this very law.

Prominent example, in this regard, is data collected for certain forms[137] of profiling, [138] which will be employed for various purposes including for behavioral targeting.[139] Another related example is the debate around the 'identifying' capacity of Internet Protocol (IP) addresses.[140] In all these cases the processing of data that relate to and target, or even single out individuals, may be involved whilst it may not *necessarily* be considered *identifiable*, thus, personal. The following paragraphs strive to show how that happens in each of these cases.

---

[136] Today's big data controllers such as Google usually argue towards strict definition of the concept of personal data although much of the '*travaux preparatoires'* leading to the DPD, for instance, tells us that wider definition and high level of protection is intended. (See, for example, *Google's Public Policy Blog* (Link) trying to define 'identifiablity' as knowing '*who* the person behind a computer is'

[137] Some aspects of profiling are caught by data protection laws: as far as they process 'data relating to an *identifiable* person.'

[138] Profiling can be defined as a process of processing and analyzing data about individuals in search of patterns, sequences or relationships to generate a profile based on which those individuals will be treated in a certain way. (See, Bygrave (2002), p. 301-02)

[139] Behavioral targeting, essentially, is a marketing tool that involves tracking people's behavior for tailored advertising. While profiling is often employed for wider purposes including in health care, insurance and law enforcement, behavioral targeting, is just one of the areas of applications of profiles, usually online.

[140] IP addresses are, essentially, unique string of numbers assigned to every device connected to the internet for to be recognized for communication purposes.

Generally, profiling is employed to extract usable information from huge amounts of data that exceed human capabilities of consideration by using powerful data mining technologies.[141] The resulting profile can be used for several of purposes including individualized marketing. Profilers use various mix of technologies to collect data from which profiles will be generated. The technological mechanisms for online tracking, for instance, continue to be more sophisticated for an average internet user to detect their existence and avoid their operation. They range from cookies and JavaScript, to less detectable Super Cookies and Ever Cookies, to location tracking and online social media tracking.[142]

Any data gathered through these methods can be subject to profiling as far as it is useful, in the mind of the profiler, for the purposes intended to be achieved. These data may/ may not fall within the scope of data protection laws depending on their capacity to identify. The danger, here, is when some of these data while particularly relate to and can single out a person—such as a profile generated from click stream data or an IP address based tracking—yet, arguably, falls outside the protection offered through data protection laws. This happens when profiling processes employ data that can only be linked to machines or other non human objects.[143]

As far as IP addresses are concerned, privacy advocates including the Article 29 Working Party have found these addresses to constitute personal data.[144] European DPAs have also taken a similar position.[145] But courts, especially in the UK where the Data Protection Act defines 'personal data'[146] differently from the DPD, have found IP addresses not to

[141] Francesca Bosco et el.,(2015), p. 4

[142] See for details, Skouma and Léonard (2015), p. 38 – 44

[143] Bygrave (2002), p. 315

[144] A29WP, Opinion 4/2007, p.16

[145] Bygrave (2014), p. 138

[146] The Act defines Personal data as "data which relate to a living individual who can be identified— (a) from those data, or (b) from those data and other information which is *in the possession of, or is likely to come into the possession of, the data controller*."(emphasis added) Thus, the legally relevant agent of identification is not 'every one' as is in the Directive.

constitute personal data.[147] The CJEU has, on the other hand, found that IP addresses are protected personal data, for instance, in its *SABAM v Scarlet extended SA* decision in 2011.[148] Yet, the court's main focus in this judgment was to rule on whether Internet Service Providers can be forced to install deep packet inspection mechanisms for protection of intellectual property. The coverage of IP addresses only figured as a secondary issue, thus, is yet to be clearly settled.[149]

Those who argue against the 'identifying capacity' of these 'data in grey area'[150] usually base their claim on the fact these data only identify machines not data subjects.[151] Even if they relate to data subjects, the argument goes, that identification happens after considerable effort and use of resources which will not satisfy the 'the existence of a *means likely reasonably to be used*' test employed by DPD.

However, much of these arguments are not defensible because in today's age of ubiquitous information auxiliary data that can be associated with these data to identify data subjects that can be found fairly easily.[152] Yet, even if we dismiss the above arguments as indefensible companies may not need to identify people through nominative data. As has been shown above, they simply target individuals based on profiles generated from data

[147] *EMI Records and Others v Eircom Ltd* [2010], paragraph 25

[148] Case C-70/10 *Scarlet Extended v SABAM* [2011], paragraph 51 of the decision specifically reads: "…the contested filtering system would involve a systematic analysis of all content and the collection and identification of users' IP addresses from which unlawful content on the network is sent. Those addresses are protected personal data because they allow those users to be precisely identified."

[149] On 28 October 2014, the German Federal Court (Bundesgerichtshof) referred a question directly about the status of dynamic IP addresses to the Court of Justice of the European Union (Case C-582/14). Specifically, the referring court is seeking to clarify whether a dynamic IP address constitutes personal data if the IP address itself is stored by an Internet service provider (ISP) while the information required to identify the user based on this IP address is held by a third party. The answer to this question will hopefully settle most of the dust around the status of IP addresses.

[150] By 'data in grey area' I am referring to those data that may not be outright classified as 'personal' yet continue to single out and target individuals.

[151] Google, for instance, generally holds this view. (See, *Google's Public Policy Blog* (Link))

[152] See, Ohm (2010), p. 1701-77

gathered that relates to them which are used, for example, for online advertisement purposes.[153] Thus, identifiability as a vital criterion that determines whether data is personal seems to inevitably fail to serve its objective of protecting privacy and related interests.

While the controversy about 'data in gray area' is yet to be settled, the damage caused to our fundamental rights from processing of these data persists. To mention few: (a) when individuals feel that their behavior is being recorded and stored somewhere in 'the sky' they will not feel free to surf through the web as would be expected for their personal autonomous development or for steady functioning of others general interests such as e-commerce. This, in turn, adversely affects over-all participation and contribution of citizens in a democratic society.[154] These problems are especially serious with non- abstract profiling that specifically target individuals: the one's Professor Bygrave calls 'specific profiling.'[155] Data subjects need to be assured that these data are guarded by principles of data protection law. (b) Based on the unnamed profiles generated from data mining and analysis, companies classify individuals in to different groups which could lead to discrimination along those lines.

Despite high expectations, the General Data Protection Regulation is not entirely clear on the issue of IP addresses and other online identifiers. It recognizes their ability to identify, and be considered as personal data, only when they are combined with other information or unique identifiers.[156] While theoretically understandable, the position taken by the regulation fails to take in to account the real context in which these data are processed. If we take IP addresses for instance most websites, like Google, never store IP addresses devoid of context; instead, they store them connected to identity or behavior.[157] Google probably knows from its log files, for example, that an IP address was used to access a

[153] See, Schwartz & Solove (2011), p. 1848- 62; Borgesius (2016), p. 256-71

[154] See, Borgesius (2016), p. 267-68

[155] Bygrave (2002), p. 303

[156] Recital 24 in the preamble to GDPR (Compromise text), recital 30 of the final version

[157]Ohm (2010), p. 1773

particular email or calendar account, edit a particular word processing document, or send particular search queries to its search engine the analysis of which can help it draw some very accurate conclusions about the person linked to any particular IP address.[158] However, even theoretically, unique online identifiers such as 'cookie identifiers' may be more reliable than the usual normative identifiers. A29WP has also, generally, made a similar observation on recital 24 of the draft regulation:

> *"…However, the Working Party considers that Recital 24, as proposed by the European Parliament and by the Council of the EU, is not satisfactory as it could be interpreted in a way that identification numbers, location data, online identifiers or other specific factors will not be necessarily considered as personal data. This could lead to an unduly restrictive interpretation of the notion of personal data."*[159]

2.1.2 Group Interests

The discussion of group interests under this sub-title is not limited to "data that targets individuals, yet may not be qualified as 'personal' for the purpose of data protection laws." This is because while data protection interests of groups may rightly arise when processing data need not necessarily 'identify' individuals—such as when groups are created by analytics[160], it can also arise when data 'identifiably' relate to group of individuals—such as in genetic data. The analysis of both is in order.

2.1.2.1 Groups Created by 'Analytics'

In these big data era new technologies and powerful analytics make it possible to collect and analyze huge amounts of data to identify patterns in the behavior of groups of individuals.[161] Data analysts are using big data to find out our shopping preferences, health status, sleep cycles, moving patterns, online consumption, friendships, etc and, in only a

[158]Ibid

[159] Article 29 Working Party 'Core topics in the view of trilogue', (17 June 2015), p.5, available at: (Link)

[160] By analytics I am generally referring to the use of information technology to exploit statistics, algorithms, and other tools of mathematics to collect and analyze large amounts of data in order to identify patterns in the behavior of groups

[161] Mantelero (2016), p. 8

few cases, mostly in intelligence circles, this information is individualized.[162] Thus, they are not necessarily 'personal' as they may not 'identify' a particular person. Even though it may not qualify as such for the purposes of data protection laws, the sheer volume of data involved in the big data world is not only highly invasive of one's privacy, but can also establish random connections based on incidental co-occurrences. In other words, big data makes the likelihood of random findings bigger.[163]

While the concept of group privacy is not new to the discourse in data protection, groups set up by big data differ from the traditional concept of a group because it involves distinctly greater predictive capacities as it uses hundreds of different variables than a few standard ones (such as sex, age, family income, marital status, place of residence) used for old profiling and categorization. Moreover, members of groups created by analytics are not aware of their membership and the consequences following from it.[164] Consequently, different privacy issues are involved in this situation than in individual or traditional group privacy issues. That, in turn, requires examination of collective interests of persons whose data is collected and analyzed to put them is a particular group.

Once large data is analyzed using powerful algorisms, patterns are drawn and people are grouped accordingly, it is no longer individual identifiability that adversely affects members of a group, it's rather, a new found group identity. In this regard, Alessandro Mantelero rightly argues that a new layer that recognizes rights of group of individuals to 'collective privacy and data protection' that is not exclusively based on atomistic individual rights should be in place.[165]

2.1.2.2 Other Forms of Non-organized Groups

Data protection laws normally require individuation for data to qualify as 'personal' except in some rare cases—such as in the Finnish law which tolerates data that can only be linked

[162] Andrej Zwitter (2014), p. 4

[163] Ibid, p. 5

[164] Mantelero (2016), p. 2

[165] Ibid, p. 2-9

to ‘a family unit’ as personal.[166] It is, hence, generally assumed that data that are liable of affecting privacy and related interests of persons are those that relate to a particular identifiable person.[167]

Despite this general assumption data can identifiably relate to various categories of ‘groups’ of individuals. Groups in which data protection interests may be implicated can be formed in various ways. Based on the state of awareness of its members, it can be formed: (a) with the consent and/or knowledge of its members—for instance, as legal persons or other non-organized groups, (b) without knowledge/consent of the members—like those formed by analytics discussed in the previous sub section. The first category of ‘groups’ can, in turn, be: (i) naturally formed—such as genetic data in which interests of close biological relatives may be involved—Or (ii) made by conscious actions of the individuals involved. The latter are usually constituted in the form of legal persons. Because of space limitation and, also since they are relatively better placed to protect their interests the discussion here will not focus on legal persons.[168] The focus here will be only on non-organized groups formed naturally or by members. In the second category of ‘groups’ emphasis was made, earlier, on collective interests arising from groups created by analytics.

As indicated, in the first category ‘groups’ are formed with the consent and/or knowledge of its members. Such groups may exist naturally without any action by its members; members will simply come to learn its existence. Or it can come in to being by members’ actions thereby implicating their privacy interests in the same data.[169] In both cases whilst

[166] Bygrave (2014), p. 136

[167] That assumption of data as ‘identifying just one data subject’ can also be reinforced from the fact that the directive does not provide a mechanism of addressing conflicts that might arise as a result of extending protection to those that are indirectly affected: Taylor refers to them as ‘secondary data subjects’(Taylor (2012))

[168] For detailed analysis of evolution, rationale and legal standing of private collective entities, See, Bygrave (2002), p. 171-298

[169] This might, for instance, be by having an identifiable information with others on the same piece of document like a group photograph

data may be gathered from one particular data subject it may, nevertheless, identifiably relate to other members. This in real terms means: genetic data identifiably relates to members of a biological family, not just to the person from whom samples are gathered & a group photograph, in the same way, relates to all persons pictured.

Fortunately, there is nothing about the 'identifiability criterion' that prohibits considering other members of a group as data subjects as long as data equally identifies them by using a means *likely reasonably* to be used. A greater problem in the directive is that it has not taken a positive step towards recognizing the possible existence of such group interests, consequently, it has not provided for a mechanism in which conflicts of interest that may arise among members of a group on the same data may be addressed.

When data relates to more than one person the resulting common interest may not always be followed by a common understanding of how data should be processed. The exercise of some rights under the DPD[170] by one data subject might potentially conflict with the preferences of another in a significant way.[171] One family member—say a person from whom a biological sample is taken—might object the exercise of his/her right by another family member of access to such data. By failing to positively anticipate and legally recognize such possibility, the DPD fails to address how such conflicts of interest might be resolved.

Absence of explicit recognition of such possibility also endangers the interests of data subjects whose interests are only indirectly involved. At the same time by failing to create a mechanism in which competing interests of members should be dealt with it also leaves data controllers with unjustifiable regulatory burden as a result of introduction of myriad of diverse interests that may follow from granting rights to—say family members—on a particular data.[172]

[170]This might be, for instance, the data subject's rights under article 12 of the DPD like a right of access, a right to the rectification, erasure or blocking of personal data.

[171] Taylor (2012), p. 116

[172] Ibid, p. 117-18

Some suggest that this problem be partly addressed by members taking an initiative from article 11 of the directive itself. Taylor and Beyleveld made the following proposition:

"The scope of Article 11 can then be restricted in a way that appropriately balances the interests of ('index case'/'primary' and 'secondary') data subjects. This does rely upon Member States taking the initiative, exercising the discretion granted them under Article 13, and appropriately protecting the fundamental rights and freedoms of all concerned."[173] That means Member States can use article 13(1) which allows them to adopt legislative measures to restrict the scope of the obligations and rights provided for in Articles 6 (1), 10, 11 (1), 12 and 21 when such a restriction constitutes a necessary measure to safeguard, in particular, the protection of the data subject or of the rights and freedoms of others.[174]

2.2 Post Identification Problems

The analysis in the previous section (section 2.1) focused, though not entirely, on interests caught up by new ways of data processing in which data while invasive and harmful may, arguably, remain non identifiable for the purpose of data protection laws. However, even with data that relates to an 'identifiable' person, thus, personal without any major doubt, there are several problems surrounding the application of the criterion of identifiability. This section will turn to the major ones.

2.2.1 The Dichotomy Syndrome

The data protection regime operates under the assumption that data is either 'identifiable' or 'anonymous.' As such the law fully applies to identifiable data while giving an exemption to data which is anonymous. Part of recital 26 in the preamble to the DPD reads:

[173] Beyleveld and Mark Taylor (2008), p. 186

[174] Article 13 (1) (g) of the DPD

> "Principles of protection must apply to any information concerning an identified or identifiable person; … whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable…[175]"

Nonetheless, one of the problematic assumptions in the application of the 'identifiability criterion' in the EU data protection law happens to be this dichotomized postulation that data falls in to one of the two categories: either 'identifiable' or 'anonymous.' The assumption is problematic when the law is applied in practice because of the following two important reasons.

First, the same data can be anonymous for one controller while identifiable for another, hence, not necessarily 'either identifiable or anonymous.' This largely hinges on a difference between controllers in an exposure to the necessary interpretive framework of identification. Not all controllers are equally exposed to what is necessary to fulfill the threshold of identifiability established by the law. Thus, two controllers holding the same data will, theoretically, be subject to completely different regulatory regimes: one that gives blanket exemption and one that obliges complete compliance. Even though, recital 26 in the preamble to the DPD makes it clear that identification by *anyone* counts for the purpose of identifiability, as we will see further below, it still creates significant uncertainty for data controllers to determine when data is not identifiable to them.

Secondly, in today's age of ubiquitous information anonymous data can be re-identified. Anonymity depends on time and context. Data which is anonymous today may be identifiable in the future; data which is anonymous in one context may be identifiable in another. Paul Ohm has very ably exposed this problem of re-identification in his UCLA Law Review article of 2010. After showing how data 'anonymized' by three sophisticated data handlers[176] has been very easy re-identified, he demonstrated that: 'data can be either

[175] Recital 23 in the preamble to the proposed GDPR (compromise text) makes, more or less, similar assertion.

[176] These were the three famous cases of American on Line (AOL) data release, Massachusetts's Group Insurance Commission (GIC) release of employee 'anonymized' health records and Netflix's release of search queries (Ohm (2010), P. 1717-23))

useful or perfectly anonymous but never both.'[177] As a result, to make data anonymous while still useful, controllers usually preferred the 'release and forget' anonymisation technique which largely relies on suppression of obvious identifiers.[178] This, in turn, paves the way for easy re-identification assisted by ever rising external information as a result of development in databases which are critical to the internet economy and proliferation of social media. Thus, the dichotomized assumption taken by the law appears to be a wrong point of departure.

It is also worth considering at this point that, Ohm further claims that in the age of easy re identification as every data assists, in some way, re identification of a data subject, the EU regime ends up being applicable to virtually every data—it becomes essentially boundless.[179] While Ohm's analysis is commendable, I believe that this latter assessment downplays the significance of the directive's requirement that there should be a 'means likely reasonably to be used.' The DPD only applies to data when there is a means likely reasonably to be used to identify the data subject; the probability and difficulty of identification matters. This keeps data whose capacity to identify is remote or involves disproportionate resources at bay.

2.2.2 Identification Factor

One of the EU data protection regime's prominences in extensive coverage is that it brings up the issue of legally relevant agent of identification.[180] It is laid down that, legally decisive is not just the ability of the controller to link a person to data but any person's

[177] Ohm (2010), P. 1704

[178] For brief overview of various techniques of anonymisation and their robustness, see A29WP Opinion 5/2014.

[179] Ohm (2010), P. 1741

[180] Recital 26 in the preamble to the DPD; the same is broached by recital 23 in the preamble to the draft GDPR (compromise text)

ability to do so.[181] The implication is that an individual is identifiable if she can be identified, directly or indirectly, by *anybody*.[182]

The law, thus, treats two very different data controllers in the same way: there is no legally relevant difference between controllers that can re-identify a given data subject from those who could not. This treatment of *'unlike parties alike'* breeds the following problems in enforcing the identifiability criterion.

First, since it is *anybody's* ability to identify that is legally relevant, when anonymity is non-obvious data controllers might have a hard time determining if data is identifiable to others. There will be no problem when data is clearly identifiable to the controller concerned. There still seems to be no problem when a controller concerned cannot identify the data subject by using a means likely reasonably to be used as far as he can reasonably assume that other parties can do so. The problem is when the ability of others to identify is not obvious/ dubious to a controller. Reigning uncertainty a controller might prefer to avoid responsibility of obedience by assuming data is anonymous while it might still be identifiable to others and vice versa. All the same, Taylor rightly argues that protecting anonymized data even if association with a particular person is not anticipated using means reasonably likely to be used by the data controller *or others* is important since there is a possibility of future unexpected re-identification, or fresh association[183], especially when data is kept at an individual level.[184] But, it is doubtful if controllers, in practice, have the incentive to do so.

Secondly, while a controller is required to comply with obligations imposed by the law, even if data is anonymous to her, performance of some responsibilities presuppose her

[181] Bygrave (2014), p. 133

[182] Taylor (2012), p. 140

[183] By 'fresh associations' Taylor is making a reference to new form of connection that may be created when—for instance, information is found from data. Information derived from genetic data, for example, can result in new form of association to other family members than the previous genetic data which, under particular circumstances, only relates to the person who gave bio samples.

[184] Taylor (2012), p. 145-46

ability to identify the data subject. These are specifically relevant for the obligations of providing information to the data subject under articles 10 & 11 of the directive, respecting data subject's right of access and objection under articles 12 and 14 respectively. Performing these obligations requires identifying the data subject and only in one of these obligations (article 11) is the obligation qualified.[185] No such qualification exists for other obligations. This can be a problem in practice because, for instance, when a controller collects data directly from the data subject as anticipated by article 10, he may not recognize the data to be identifiable at the time of collection, thus, may fail to appreciate the need to provide this information. Somebody conducting surveys in the street may deliberately avoid collecting generally identifiable data precisely because he wishes to avoid the responsibilities associated with collecting 'personal data'.[186]

However, even if a controller finds a way to identify the data subject for the purposes of compliance, the imposition of these obligations have the effect of turning a merely identifiable, or even anonymous, data in to an identified one. As it encourages identification such an obligation will, thus, be counterproductive in many circumstances.[187] Therefore, treating *'unlike parties alike'* may be problematic in enforcing the criterion of 'identifiability.' These difficulties appear to be the reasons why some national laws, like the UK's Data Protection Act of 1998 preferred to go around a clear intention of the directive in limiting a legally relevant agent of identification.

On the bright side, the GDPR provides that, if the data processed by a controller do not permit the controller to identify a natural person, data controller should not be obliged to acquire additional information in order to identify the data subject for the sole purpose of complying with any provision of this Regulation.[188] However, the controller should not

[185] When data is not directly gathered from the data subject Article 11(2) excuses a controller's obligation to notify if the provision of such information proves impossible or would involve a disproportionate effort

[186] Taylor (2012), p. 143

[187] Schwartz & Solove (2011), p. 1876-77

[188] Article 10(1) and Recital 45 in the preamble to the regulation (compromise text). Article 10(2) specifically mentions the possible suspension of obligations under articles 15-18 of the regulation.

refuse to take additional information provided by the data subject in order to support the exercise of his or her rights.[189] Paragraph 2 of Article 10, though, is not entirely clear. The first sentence still seems to sustain the controller's obligation to inform the data subject, while at the same time acknowledging the controller not being in a position to identify the data subject. The final version of the regulation did not clarify the confusion.[190]

[189] Article 10(2) and Recital 45 in the preamble to the regulation (compromise text)

[190] Article 11(2) of the GDPR

Chapter 4

Conclusion and Recommendations

4.1. Conclusion

The analysis in this study is made in an endeavor to challenge the conceptual predispositions behind the two building blocks of the definition of personal data under the current and *en route* EU data protection rules. These two building blocks in that definition are: '*information'* capable of '*identifying'* the data subject.

The terms data and information, though key legal jargons, are often taken for granted and insufficiently, if at all, defined in data protection discourse. As technology, particularly in the field of bio technology develops, however, a workable definition is increasingly needed because the blurring of the boundary between human body and technology may trigger application of laws intended for the informational world—such as data protection—to the biological world.

A close look at the Data Protection Directive, in this regard, reveals the absence of a positive intention by the architects of the directive to consider biological materials as data/information. This can be observed from the way personal data itself is defined, the scope of application of the directive is crafted, and from the semantic inhospitality—to biological materials—of some of the key terminologies employed throughout the directive.

While it makes mention of 'biological materials,' it does not appear that the General Data Protection Regulation is intended to be applicable to such materials. Whilst recital 26 in the preamble to the regulation tempts us to consider biological materials as information, the syntax used in the recital is ambiguous. Moreover, recital 35 in the preamble of a final version of regulation (published on EU Official Journal on 4 May 2016) indicates that the tempting syntax in recital 26 of the compromise text is a result of poor draftsmanship. Furthermore, other structural problems in the directive such as scope, and usage of key words persist in the regulation as well.

The DPD and its preparatory materials indicate that the architects did not have the issue of biological materials on the table. The same assumption, however, can't be made about the General Data Protection Regulation as it introduces numerous tempting terminologies. By introducing proper terminologies such as—biological materials and genetic data—the architects of the regulation tried to create an appearance that the regulation applies to biological materials without providing any real substance in this regard.

The question of whether biological materials *should be* treated as personal data is far from consensus. Scholars who pay more attention to pragmatic considerations have forwarded the view that biological materials should be regarded as personal data/information. Other scholars, commentators and data protection enforcement authorities have opposed this view mainly based on conceptual logic, arguing that data is a formalized representation of objects while information comprises cognitive elements involving comprehension of that representation.[191]

However, a range of developments in molecular biology and nano-technology, largely mediated by advances in ICT, are at odds with the conceptual distinction between data and information. First, proteins—which make up the basis for almost everything in the human body—are made as per 'the information' obtained by reading the order of strands of nucleotides in our DNA. Thus, information lies at the very origin of life. Secondly, ambitious scientific initiatives such as the BRAIN (Brain Research through Advancing Innovative Neuroethologies)—which intends to decode neurons in our brain much like the Human Genome Project did for our genome—may lead to our continued existence as information. Thirdly, the ongoing human enhancement projects (HEP) are clouding the distinction between the human body and technology. And finally, proliferation of bio banks and the increasing deployment of biometric technologies are converting aspects of our bodies in to processable digital data.

In addition, multiple pragmatic considerations beseech the collapse of the distinction between data, as carrier, and information, as a result of processing data. First, it is difficult

[191] Bygrave (2015), p. 6-7

to find distinguishable interpretive potential between data and information; it all turns on availability of the right interpretive framework. Secondly, the lacunae in the regime governing bio banks might be assisted by the more comprehensive rules under data protection which also possesses better enforcement mechanisms. And finally, considering biological materials only as a medium may, sometimes jeopardize our fundamental rights even more, thus, making the maintenance of the distinction morally indefensible.

Despite a crumbling conceptual rigor that distinguishes human biological materials from data/information and various pragmatic considerations that increasingly challenge such distinction, collapsing differences that have been maintained in the regulatory discourse for such a long time is not without its own drawbacks. First, it will overstretch the rules that are already said to have a long arm which may be counterproductive for their effective enforcement. Secondly, while 'consent' enjoys a relatively central role under the directive when with regard to sensitive data—the category to which biological materials would most probably belong—it is doubtful that consent plays or would play a central role in the processing of personal data in general. As consent remains central in other laws traditionally concerned with human biological materials the extension of the DPD or the GDPR to biological materials only poorly meets the normative position of consent maintained by these laws. Finally, extending biological materials to the data protection regime would demand DPAs to have more financial and human resources with the requisite skills to handle the peculiarities of biological materials.

The application of the other essential constituent in defining personal data under EU data protection laws i.e., *identifiability*, is also confronted by multiple challenges. This analysis is made by dividing those problems in to, first, pre identification problems to capture how technological developments have allowed different levels of risk to our fundamental rights by processing data without a need to *'identify'* individuals, and secondly, post identification interests to capture problems with the criterion even when it applies to data that clearly identifies.

It is shown that the criterion of '*identifiability*' under the DPD fails to make data personal and protect our fundamental rights in this big data world, when practices such as profiling

turn on processing data that is entirely non identifiable yet can target or single out individuals. Despite high expectations, the General Data Protection Regulation is also not entirely clear on the issue of IP addresses and other online identifiers. It recognizes their ability to identify, and to be considered as personal data, only when combined with other information or unique identifiers. In addition, the criterion fails to adequately serve the interests of people in a group: whether the group is created without their knowledge—such as those made by Analytics, or those that are made by/known to them.

The application of the criterion to data that 'identifies' is also encircled by many challenges. One such difficulty stems from a dichotomized assumption made by the law which suggests that data is either 'identifiable' or 'anonymous.' Despite this assumption, first, the same data can be anonymous to one controller while identifiable for another. Secondly, given the ubiquity of information anonymous is increasingly identifiable with relative ease. Coupled with the problems faced by the 'legally relevant identification factor' these difficulties may create significant uncertainty, in some situations, as to whether data is identifiable and thus a proper subject of data protection laws.

4.2. General Recommendations

The application of data protection laws to human biological materials would clarify the ambiguity created by unclear usage of the terms data and information in the current definition. As analyzed in this study, many developments in sciences and a range of pragmatic evidences also suggest that these materials should be considered as data or information. However, doing so requires the introduction of some basic changes to these laws. Changes are required on how we craft the scope of application of these laws, our usage of key phrases throughout the legislations and how we define the concept of personal data itself. It also calls for an independent and thorough study in to the regime governing biological materials to determine whether data protection laws can be the sole regulator in this area.

The criterion of identifiability should be approached with judicial activism which aims to sustain the original objective of data protection laws by broadly interpreting the criterion.

Approaches such as defining identifiability to mean only 'nominative data' should be avoided because such approaches provide a loophole through which data that clearly jeopardize our fundamental rights and central values may remain non-personal. Though not clearly stipulated by the GDPR itself, mentions of new identifiers such as *online identifiers* should be taken as a green light by courts to broadly draw from the criterion of identifiability. The phrase 'singling out' used in recital 23 of the GDPR can, in this regard, be taken as a positive step to improve the criterion of identifiability to at least cover profiling practices that are 'specific.'

Finally, it is fair for one to argue that these recommendations simply stretch the boundary of a law which is already accused of lacking one, at least a clear boundary. While a valid point, one should also consider, first, as far as privacy is concerned, it is public life that extended to the private realm which is fueled by developments in Information and Communications Technology. Technology came to our doors and exposed our private lives, which created fear and a sense of powerlessness that must be addressed. Finally, especially for those who argue that data protection should be limited to data about our private lives,[192] it should be clear that it is for a reason that 'the right to data protection' is made to separately exist from the right to 'privacy' by the Treaty of the Functioning of the European Union[193] and the Charter of Fundamental Rights of the European Union.[194] By so doing, the right to data protection aims to breed fairness in the processing of data about persons and is not limited to protection of their private lives. The same is now recognized by the GDPR[195] when it provides, under article 4(2) 'protection of fundamental right especially the right to data protection'—not privacy—as was the case in the corresponding provision of the DPD.[196]

[192] See, Borgesius (2016), p. 270, and the references cited therein

[193] Article 16

[194] Article 8

[195] Article 1(2) of the GDPD (Regulation 2016/679); See also, Costa and Poullet (2012), p. 255

[196] Article 1(1) of the DPD

Bibliography

1. List of Judgments/ Decisions

CILFIT v Ministery of Health, Court of Justice of the European Union, Case- 283/ 81 [1982]

EMI Records and Others v Eircom Ltd [2010] IEHC 108, Irish High court, 16th April, 2010

L, Regina (on application of) v South Yorkshire Police (Consolidated Appeals) [2004] UK House of Lords 39; 1 WLR 3223

*M.S. v Sweden***,** the European Court of Human Rights, Strasbourg, 25 February 1997

Patrick Breyer v Bundesrepublik Deutschland, Court of Justice of the European Union, C-582/14, lodged on 17 December 2014.

S and Marper v United Kingdom, European Court of Human Rights, (App no 30562/04 and 30566/04, 4 December 2008

Scarlet Extended v SABAM, Court of Justice of the European Union, C-70/10 [2011]

Z Vs Finland, the European Court of Human Rights, Strasbourg, 27 August 1997

2. Treaties/ Statutes

Charter of Fundamental Rights of the European Union (2012/C 326/02), of the European Parliament, the Council and the Commission

Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), Rome 1950

Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data CETS No.: 108, Treaty open for signature by the member States and for

accession by non-member states, opened for signature at Strasbourg on 28/1/1981, entered in to force on 1/10/1985

Council of Europe, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, European Treaty Series - No. 164, Oviedo, 4.IV.1997

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

Directive 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.

European Commission - Press release Agreement on Commission's EU data protection reform will boost Digital Single Market, Brussels, 15th of December 2015

Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), January 2012.

Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), Brussels, 25.1.2012 COM(2012) 11 final 2012/0011 (COD)

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)

UNESCO, International Declaration on Human Genetic Data, 16 October 2003

World Medical Association (WMA), 2008, World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964, amended in 1975, 1983, 1989, 1996, 2000, 2002, 2004, and 2008.

3. Secondary Literature

Amos, Martyn. *Theoretical and Experimental DNA Computation*, Natural Computing Series, Springer, 2005

Article 29 Working Party, "Core topics in the view of Trilogue" (17 June 2015), available at (Link)

Article 29 Working Party, "Opinion 05/2014 on Anonymisation Techniques," Adopted on 10 April 2014, 0829/14/EN WP216

Article 29 Working Party, "Opinion 4/2007 on the Concept of Personal Data," Adopted on 20th June, 2007, 01248/07/ENWP 136

Australian Law Reform Commission (ALRC) and Australian Health Ethics Committee (AHEC), "Essentially Yours: The Protection of Human Genetic Information in Australia," Report No 96 (ALRC/AHEC, Sydney 2003

Berg, Jeremy, John Tymoczko and Lubert Stryer. *Biochemistry*. New York: W H Freeman, 2002

Beyleveld, Deryck and Mark Taylor, "Patents for biotechnology and the data protection of biological samples and shared data," in *The Protection of Medical Data: Challenges of the 21st Century*, edited by Jean Herveg, 127–48. Louvain-la-Neuve: Anthemis, 2008

Beyleveld, Deryck *et al.*, "The UK's Implementation of Directive 95/46/EC" in, *Implementation of the Data Protection Directive in Relation to Medical Research in Europe.* Edited by Deryck Beyleveld *et al.* Ashgate, 2004

Bird & Bird, "EU Framework Revision: Overview," available at: (Link)

Bosco, Francesca e*t al*., "Profiling Technologies and Fundamental Rights and Values: Regulatory Challenges and Perspectives from European Data Protection Authorities," in *Reforming European Data Protection Law*, edited by Serge Gutwirth *et al.* Springer, 2015

Bygrave, Lee. "Information Concepts in Law: Generic Dreams and Definitional Daylight," *Oxford Journal of Legal Studies*, Vol. 35, No.1 (2015): 1-30

Bygrave, Lee. "The Body as Data? Bio-bank Regulation via the "Back Door" of Data Protection Law," *Law, Innovation and Technology* Vol. 2, issue1 (2010): 1-25

Bygrave, Lee. *Data Privacy Law, an International Perspective*, Oxford University Press, Oxford, 2014

Bygrave, Lee. Data Protection Law: Approaching Its Rationale, Logic and Limit, Kluwer Law International, The Hague, 2002

Calladine, Chris, Horace Drew, Ben Luisi and Andrew Travers, *Understanding DNA: The Molecule and how it Works.* London: Elsevier Academic Press, 2004

Costa, Luiz and Yves Poullet, "Privacy and the regulation of 2012," *Computer Law and Security Review*, Vol. 28, Issue 3 (2012): 254-262

EC Commission, "Commission's commentary with respect to its amended proposal for a data protection Directive of15 October 1992," COM (92) 422 final—SYN 287

EU Commission's Directorate-General for Research and Innovation Science in society, *Bio-banks for Europe: A challenge for governance*, Report of the Expert Group on Dealing with Ethical and Regulatory Challenges of International Bio-bank Research, 2012

European Union Agency for Fundamental Rights, *Handbook on European data protection law*, Belgium, Council of Europe, 2014.

Frederik, Zuiderveen Borgesius, "Singling out people without knowing their names – Behavioral targeting, pseudonymous data, and the new Data Protection Regulation," *Computer Law & Security Review*, Vol. 32, Issue 2 (2016): 256-271

Georgia Skouma and Laura Léonard, "On-line Behavioral Tracking: What May Change after the Legal Reform on Personal Data Protection," in *Reforming European Data Protection Law*, edited by Serge Gutwirth et al., Springer, 2015

Isabelle Abbey, "News and Views: The Brain Activity Mapping Project – What's the plan?" April 24, 2013

John Steward, "Human Enhancement," *Dartmouth Journal of Undergraduate Science*, in Fall 2013.

Kaku, Michio. The Future of the Mind: The Scientific Quest to Understand, Enhance and Empower the Mind. New York: Doubleday, 2014

Koops, B.J. "the trouble with European data protection law," *International Data Privacy Law*, Vol. 4, No. 4 (2014): 250-261

Mantelero, Alessandro, "Personal data for decisional purposes in the age of analytics: From an individual to a collective dimension of data protection," *Computer Law & Security Review:* Vol. 32, Issue 2 (2016): 238-255

Meglena Kuneva (European Consumer Commissioner (Jan. 2007- Feb. 2010), *Roundtable on Online Data Collection, Targeting and Profiling Brussels*, *Keynote Speech* (SPEECH/09/156), 31 March 2009

Moerel, Lokke. "The long arm of EU data protection law: Does the Data Protection Directive apply to processing of personal data of EU citizens by websites worldwide*?"* *International Data Privacy Law*, Vol. 1, No. 1 (2011): 28-46

Murnaghan, Ian, "the Importance of DNA," *Explore DNA*, Feb. 2016, Available at: (Link)

Ohm, Paul. "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymisation," *UCLA Law Review*, Vol. 57, (2010):1701- 1777

Perry Rotella, "Is Data The New Oil?" *Forbes*, April 2, 2012

Ploeg, Irma van der, "Biometrics and the Body as Information: Normative Issues of the Socio- Technical Coding of the Body," in *Surveillance as Social Sorting: Privacy, Risk, and Automated Discrimination*, edited by David Lyon, Routledge, 2002

Ploeg, Irma van der, "Genetics, biometrics and the informatization of the body," *Ann 1st Super Sanità*, Vol. 43, No. 1, (2007): 44-50

Reed, Chris, *Making Laws for Cyberspace*. Oxford: Oxford University Press, 2012

Ridley, Matt. *Genome: The Autobiography of a Species in 23 Chapters*, New York: HarperCollins Publishers, 1999

Schwartz, Paul & Daniel J. Solove, "the PII Problem: Privacy and a new Concept of Personally Identifiable Information," *New York University Law Review*, Vol. 86, (2011): 1814-1894

Science Magazine, "Injectable, Cellular-Scale Optoelectronics with Applications for Wireless Optogenetics," 12 April 2013

Taylor, Mark. Genetic Data and the Law: A Critical Perspective on Privacy Protection, Cambridge University Press, 2012

The Guardian: "Yes, nanoscience can enhance humans – but ethical guidelines must be agreed," Monday, 3 June 2013

The Stockholm Programme — an open and secure Europe serving and protecting citizens, OJ C 115, 4.5.2010.

Wendy Schrama, "How to Carry out Interdisciplinary Legal Research some Experiences with an Interdisciplinary Research Method," *the Utrecht Law Review*, Volume 7, Issue 1 (2011): 147-162

World Economic Forum, *Personal Data: the Emergence of New Asset Class*, An initiative of World Economic Forum in collaboration with Brain and Company, Inc., January 2011.

Zwitter, Andrej, "Big Data ethics," *Big Data & Society*, July–December 2014: 1–6