THE BOUNDARIES OF PERSONAL DATA: INSIGHTS FROM THE UK AND EU APPROACHES

Tee May Ern*

Introduction

In today's digital era, personal data has garnered a reputation as the new 'oil' due to its immense value in driving and influencing the way many businesses operate.[1] Commonly, it serves as crucial input for the purposes of targeted advertising. Universities, for example, would want their online advertisements directed at prospective students, and it is only by leveraging on the personal data of Internet users (e.g. age, location) that they can effectively ensure that their promotional message are directed at the right individuals, rather than indiscriminatorily to the public. Similarly, a music streaming platform may, for example, leverage on personal data such as a user's listening history and favorite genres to curate customized playlists that align with each user's unique taste in order to enhance the user's experience, which ultimately drives user satisfaction and loyalty.

Such utilisation and processing of personal data is made possible through technological advancements, which has made it increasingly easier and cheaper to store, transmit and access vast amounts of personal data. However, this rapid advancement in data collection and dissemination has also brought about significant concerns regarding the protection of personal data. Unauthorized disclosures of personal data, particularly sensitive personal data e.g. bank details or passwords, can have devastating consequences for individuals, ranging from financial losses to enduring distress and worry.

Positively, as a response to these growing threats, legislators globally have in recent times taken proactive measures to establish robust data protection frameworks.[2] However, it is important to note that these frameworks usually apply specifically to information classified as "personal data" only. Put simply, it is only where the information is classed as "personal data" will that information fall within the purview of data protection frameworks. The definition and, perhaps

* Lecturer, Faculty of Law and Government, HELP University

[1] The world's most valuable resource is no longer oil, but data (*The Economist*, 6 May 2017) <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data> last accessed 2 July 2023

[2] The General Data Protection Regulation (GDPR), which saw strong measures taken by EU countries to safeguard personal data, came into effect in 2018. Another prominent data protection law that was also introduced in 2018 is the California Consumer Privacy Act 2018. The trend of introducing and strengthening data privacy is not confined only to Western countries, notable efforts have had also been taken in the Asian countries. Singapore, for example, has made recent amendments (in 2020) to its existing Personal Data Protection Act 2012 to increase the enforcement of its data protection regime. In 2022, Thailand's Personal Data Protection Act 2019 came into effect, representing their first legislation specifically targeted at protecting personal data. Additionally, India is anticipating a new legislation on data privacy with its proposed Digital Personal Data Protection Bill 2022.

more importantly, interpretation of personal data is thus crucial in determining the boundaries and scope of data protection laws.

Currently, although many countries have legislation which defines "personal data", there remains a notable lack of clarity regarding the precise scope and meaning of this term, which raises questions on what is and is not covered by such laws. While some information is obviously considered personal data (e.g. name, address, phone number), there exist many instances where the distinction is less clear. These grey areas or borderline cases include information like handwritten notes, a compilation of someone's favourite movies, or even a queue number. Unfortunately, there have been few test cases that specifically address the interpretation of personal data. Most existing cases raising data protection issues tend to focus only on information that clearly fits the common-sense notion of personal data.

This essay aims to provide a comparative analysis on the interpretation of personal data by the United Kingdom (UK) and European Union (EU) courts. These jurisdictions were deliberately chosen not only because the definitional elements of "personal data" in many countries closely mirror those used in the UK and EU, but also because these two jurisdictions present different approaches to the interpretation of personal data. Specifically, the EU adopts a very expansive approach to the interpretation of "personal data", while the UK adopts a slightly narrower one. In examining the differing approaches, this essay further aims to evaluate the implications of adopting an expansive interpretation of "personal data", and to analyse whether such an approach is inherently advantageous.

Definition of "personal data"

The General Data Protection Regulation (GDPR) is an EU regulation which lays down data protection rules and measures which all EU states are bound by. It is a very influential piece of regulation for two main reasons. First, it represents a modern and robust framework targeted specifically at the protection of personal data, and often considered as the gold standard in the context of data protection. Secondly, it carries an extraterritorial effect which necessitates compliance of many companies and organizations operating outside the EU.[3] As such, many jurisdictions have created or amended their data protection laws to closely align and mirror the principles of the GDPR.

The GDPR defines "personal data" as "any information relating to an identified or identifiable natural person".[4]

In the UK, the Data Protection Act 2018 implements the GDPR and adopts a virtually identical definition for "personal data" as "any information relating to an identified or identifiable living

[3] Article 3 of the GDPR provides the applicability of the GDPR to non-EU data controllers or processors, so long as their activities include offering goods or services, irrespective of whether payment is required, to individuals in the EU, or the monitoring of behaviour of individuals in the EU.
[4] Article 4 of the GDPR

individual".[5] This alignment comes as no surprise, considering that the UK Act was created based on the GDPR.

Interestingly, there exist a remarkable consensus in how "personal data" is defined across numerous jurisdictions, with many adopting a similar understanding and definition of the term in their data protection framework. In Malaysia, for example, the Personal Data Protection Act 2010 defines "personal data" as "[a]ny information in respect of a commercial transaction which … relates directly or indirectly to a data subject, who is identified or identifiable from that information or from that and other information in the possession of a data user …".[6] In Singapore, the Personal Data Protection Act 2012 states that "personal data" means "data, whether true or not, about an individual who can be identified (a) from that data; or (b) from that data and other information to which the organisation has or is likely to have access".[7] As a final example, in Hong Kong, the Personal Data (Privacy) Ordinance defines "personal data" to mean "any data – (a) relating directly or indirectly to a living individual; (b) from which it is practicable for the identity of the individual to be directly or indirectly ascertained …".[8]

The definitions of "personal data" across the aforementioned jurisdictions noticeably incorporates several common essential elements, namely, "information" (or "data") that is "relating to" (or "about") an "identified" or "identifiable" person/individual.

An analysis on how the EU and UK courts interpret the definitional elements of "personal data" can thus be highly valuable for many other jurisdictions to gain insights into the different interpretive approaches and their respective implications.

Among the said common elements, some are widely understood and applied similarly. For instance, the concept of "information" or "data" is generally accepted to encompass both true and untrue, accurate and inaccurate, and objective and subjective information. Accordingly, an inaccurate profile about an individual can fall within the meaning of "personal data". Moreover, there is general consensus that such information can come in any form, ranging from a handwritten note to a painting or picture. Similarly, the notion of a "person" or "individual" is commonly interpreted as pertaining exclusively to living individuals.

What is perhaps less clear are the elements of "relating to" and "identified or identifiable", which this article will now turn to analyse.

"Relating to"

European Union

[5] Data Protection Act 2018, S3(2)
[6] Personal Data Protection Act 2010, S4
[7] Personal Data Protection Act 2012, s2
[8] Personal Data (Privacy) Ordinance, S2(1)

The case of *Nowak v Data Protection Commissioner*[9] represents one of the most recent interpretations by the Court of Justice of the European Union (CJEU) on the concept of "relating to" in the context of personal data. In this case, Mr Nowak had taken some accountancy exams and was dissatisfied after failing a particular paper multiple times. After failing for the fourth time, he sought access to all his personal data held by the examination board, including his written examination scripts, as well as the comments provided by the examiner. The central issue which the CJEU was asked to consider was whether written answers submitted by a candidate at a professional examination, along with any examiner's comments, constitute "personal data". In particular, focus was placed on whether such written answers constitute information "relating to" the candidate.

The CJEU held that information "relates to" a person if "the information, by reason of its content, purpose or effect, is linked to a particular person".[10] Applying this principle to the present case, the CJEU made the following findings: -

With regards content, the CJEU held that the content of a candidate's examination answers reflects his extent of knowledge, competence, intellect, thought processes, and judgment.[11] Where the examination scripts were handwritten (as was in this case), it would also contain information as to the candidate's handwriting.[12]

Additionally, the purpose for collecting examination answers was to evaluate the candidate's professional abilities and his suitability to practice the profession concerned.[13]

Lastly, the use of a candidate's examination answers would carry an effect in determining his success or failure at the examination concerned. It is thus capable of affecting the candidate's rights and interest insofar as it may determine or influence the candidate's chances of entering his desired profession.[14]

Accordingly, the CJEU concluded that Mr Nowak's written answers did constitute his personal data as the content, purpose and effect of his examination scripts all linked to him.

Notably, the CJEU went further to say that the examiner's comments too were capable of constituting a candidate's personal data, similarly because its content, purpose or effect links to the candidate. The content of an examiner's comments link to a candidate as it represents the examiner's opinion and assessment of the candidate's performance in the examination, specifically regarding his knowledge and competence in the relevant field.[15] Its purpose is to

[9] Case C-434/16, *Nowak v Data Protection Commissioner* ECLI:EU:C:2017:994
[10] Case C-434/16, *Nowak v Data Protection Commissioner* ECLI:EU:C:2017:994, at [35]
[11] Ibid, at [37]
[12] Ibid
[13] Ibid
[14] Ibid, at [39]
[15] Ibid, at [43]

record the evaluation by the examiner of the candidate's performance, and its effect determines the candidate's future prospects (e.g. his employability).[16]

Nowak thus illustrates that a set of personal data could belong to more than one individual at the same time. In this case for example, the examiner's comments were found to be both the candidate and the examiner's personal data. This may no doubt lead to questions on how data protection laws might apply in the event of a conflict of interest between such individuals over the processing of personal data, but it is beyond the scope of this article to assess that issue.

United Kingdom

In the UK, the leading case which sheds light on the meaning of "relating to" is *Durant v Financial Services Authority (FSA)*[17]. In this case, Mr. Durant had made an access request to the FSA, requesting for it to provide him with his personal data which the FSA held electronically and in manual files. The FSA responded to the request by providing Mr Durant with some information which it held electronically, but refused to provide any information held on their manual files on grounds that such information did not constitute his "personal data". An issue which the court had to consider was whether or not documents concerning an individual could be classified as his personal data. This was decided in the negative at trial, and on appeal to the Court of Appeal, it was again so decided.

The Court of Appeal decided that a narrow interpretation should be given to the term "personal data". Importantly, it held that the simple mentioning of an individual in a document does not automatically make it that individual's "personal data".[18] In order to constitute personal data, the information must have the individual as its focus.[19] The Court of Appeal explained that for information to qualify as "relating to" an individual, it must be information that "affects [the individual's] privacy, whether in his personal or family life, business or professional capacity".[20] In determining this, two factors are of particular assistance. First, was the information of biographical significance in the sense that it goes beyond recording the individual's involvement in a matter or event without personal connotations. Second, did the information carry individual focus in the sense that it primarily focused on the individual rather than some other person with whom they may have been associated or had an interest.[21]

In this case, the information which Mr Durant was seeking access to were documents from an FSA investigation into a complaint which Mr Durant had made against a Bank. The Court of Appeal held that this did not fall within Mr Durant's "personal data", observing that although Mr Durant had initiated the complaint in question, it did not follow that all information relating to those complaints could be considered his personal data.

[16] Case C-434/16, *Nowak v Data Protection Commissioner* ECLI:EU:C:2017:994, at [43]
[17] [2003] EWCA Civ 1746
[18] *Durant v Financial Services Authority* [2003] EWCA Civ 1746, at [28]
[19] Ibid
[20] Ibid
[21] *Durant v Financial Services Authority* [2003] EWCA Civ 1746, at [28]

"Identified or identifiable"

European Union

The GDPR explains that an identifiable person is "one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".[22]

In essence, this criterion acknowledges that personal data includes not only data that directly reveals someone's identity but also data that enables the identification of an individual. It thus encompasses situations where an organisation can identify a person based on a set of data, or based on a combination of that data with other data, which it holds. Take for example, a scenario where there is data retrieved from an anonymous survey about an employee working in a particular multinational company. On its own, this information may not distinguish one employee from many others. However, when combined with a separate dataset containing additional details on the individual's age, gender, hometown, department which he/she works in, number of years he/she has been working in the company, it becomes increasingly likely to identify a specific individual matching these combined characteristics. This will thus make the initial data one which enables the identification of an individual.

An aspect which is very much open to interpretation is whether an organisation must already have access to the additional data required for identification, or whether it also cover situations where such data needs to be acquired from third parties. Take for example, a scenario where an organisation (Party A) holds a customer's grocery store receipt. This alone may not enable that customer to be identified, assuming that it does not know who the receipt belongs to. However, when this receipt is combined with information such as the timing and location of the purchase, it can enable the identification of the customer by cross-referencing it with CCTV footage held by the grocery store (Party B). In this case, the combination of the receipt data and the CCTV footage, may collectively lead to Party A being able to identify the customer. Would this then make the receipt data on its own "personal data" on the basis that identification is possible when the receipt data is combined with other data, even if this other data is held by a separate party? If the answer is in the affirmative, then should we also take into account other factors such as the likelihood, feasibility and lawfulness of obtaining the separate dataset?

This issue of what information would relate to an "identifiable" person was addressed by the CJEU in *Breyer v Bundesrepublik Deutschlan*[23]. In this case, Mr Breyer initiated legal proceedings against the German authorities due to their practice of storing IP addresses of individuals who accessed certain websites for the purpose of combating cybercrime. Mr Breyer contended that this storage of IP addresses violated data protection laws. In response, the German authorities argued that IP addresses did not qualify as "personal data".

[22] Article 4 GDPR

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

An important issue which the CJEU was asked to consider was whether IP addresses constituted information enabling the identification of an individual in circumstances where the additional data needed for such identification is held by a separate party (in this case, the user's internet service provider). The CJEU answered in the affirmative, emphasising that it is not a requirement that all the information enabling the identification of an individual be in the hands of one person.[24] In this case, the CJEU held that although an IP address itself could not lead to the identification of Mr Breyer, it nevertheless carried the potential of him being identified when said IP address was supplemented with additional data held by Mr Breyer's online service provider.[25] This was deemed sufficient to make IP addresses "personal data".

The CJEU did, however, acknowledge that two considerations must be taken into account. First, the lawfulness of obtaining the additional data; an individual would not "identifiable" if additional data needed for such identification can only be obtained illegally.[26] Second, the proportionality or reasonability in obtaining the additional data; an individual would not be "identifiable" if it requires a disproportionate time, effort and manpower to obtain such a data that the risk of identification is practically insignificant.[27]

In this case, it was acknowledged that it would not have been legally permissible under German law for the authorities to directly acquire additional information from internet service providers. However, the CJEU held that in the event of a cyberattack, the German authorities could still approach the competent authorities, such as the police and German courts, to lawfully acquire such additional data.[28] Based on this reasoning, the CJEU concluded that IP addresses constitute "personal data" since there exists a potential for lawfully obtaining additional information that could lead to identification.

United Kingdom

In the UK, the term "identifiable" has not been explicitly addressed by the courts. In *Ittihadieh v 5-11 Cheyne Gardens Rtm Company Ltd and Others*[29], an English case decided after *Breyer*, the Court of Appeal seemed to acknowledge and endorse the interpretation of "identifiable" as established by the CJEU in *Breyer*. It explained that "identifiable" means "capable [of being identified from particular data] without disproportionate effort",[30] an interpretation which aligns with the understanding derived from the CJEU's ruling in *Breyer*.

While the English courts do appear to accept the meaning of "identifiable" as applied in *Breyer*, it remains unclear whether the UK courts would approach the assessment of "proportionality" in the same rigour as the CJEU such that it includes circumstances where the additional

[24] Ibid, at [43]
[25] Ibid, at [45]
[26] Ibid, at [46]
[27] Case C-582/14, *Breyer v Bundesrepublik Deutschland*, ECLI:EU:C:2016:779, at [46]
[28] Ibid, [47]-[48]
[29] [2017] EWCA Civ 121
[30] Ibid, at [97]

information can only be obtained with the assistance from a competent authority and third party (e.g. through the court making a court order to a third party requiring disclosure of additional data to enable identification).

Evaluation on the implications of a broad interpretation

It is apparent that the EU perspective is to take an expansive approach to data protection through its liberal interpretation of the definitional elements of "personal data". Based on *Breyer*, even if the data controller holds data which is insufficient to identify an individual, such data can still be considered as said individual's "personal data" so long as is it possible to lawfully and reasonably/proportionately obtain additional data from third parties to identify the individual. In fact, application of the proportionality test by the CJEU in *Breyer* makes it clear that the CJEU is willing to accept as proportionate efforts which are quite excessive, such as obtaining a court order to compel a third party to disclose additional data needed for the identification of an individual.

Following the EU's approach, it may very well be that seemingly anonymous information too can be deemed as an individual's personal data. For example, responses to surveys, even if conducted anonymously, can be the "personal data" of the respondent in question if it contains information which can be used to identify that individual. Similarly, a comment or post from a social media account using a fake name too can be deemed as that user's "personal data" if his identity can be ascertained through his digital footprint.

The EU's broad approach is also apparent in its interpretation of the element "relating to" as seen in *Nowak*, which practically suggest that most, if not all information, that is even remotely connected to an individual can be that individual's personal data. Based on the CJEU's approach in *Nowak*, it could be argued that a student timetable, for example, could be a teacher's personal data. Because its content reflects the teacher's expertise and (a part of his/her) schedule, its purpose is to make the teacher aware of where and when he/she needs to be present for teaching activities, and non-compliance with said timetable could carry significant consequences to the teacher (e.g. disciplinary action or dismissal).

This stands in contrast to the English perspective where a slightly narrower approach was taken; information will only be deemed as the "personal data" of an individual if it intrinsically links to that individual in the sense that it is biographical or has the individual as its focus. Based on *Durant*, the example of a student timetable would unlikely be considered the teacher's "personal data" as it is not biographical, and it does not have the teacher as its focus. Rather, its focus is to provide students with a schedule of classes to attend.

A broad interpretation of "personal data" holds significant benefits in the digital age we live in, because what many people often fail to realise is that even seemingly insignificant pieces of information about individuals can, when combined with additional data, lead to their identification. In fact, a 2000 study conducted in the US showed that the combination of a ZIP

code, date of birth, and gender was enough to identify 87% of the US population.[31] Considering that such accurate identification was possible in 2000, it further becomes unimaginable the extent to which technological advancements over the last 20 years have amplified the potential for rapid and precise identification based on even simpler pieces of information about individuals.

Moreover, the combination of data, even seemingly unrevealing data about us, may disclose a substantial amount of information about ourselves which we may not be aware of. It could be envisaged, for example, that a call log indicating the timing, duration, and frequency of calls can provide insights into one's closeness in relationship with a person. By cross-referencing this data with information held by network service providers, it further becomes possible to determine that individual's location, frequented places, preferred modes of transportation, and even daily routines. As a further example, an individual's IP address alone may not reveal much about this individual. But when this data is combined additional data from the user's internet service provider and his/her cookie data, it becomes possible to acquire that individual's search history and subsequently construct a detailed profile of his/her private and professional life. With the rapid advancement of technology which has not only enabled more data to be collected, but also for these to be easily combined with other data and analysed, it certainly seems necessary and desirable to apply data protection laws broadly.

That said, an expansive application of data protection laws is certainly not without problems. A concern raised by Purtova with the EU's expansive approach is that by including a wide range of information under "personal data", data protection laws will easily become "the law of everything".[32] Purtova goes so far to demonstrate how based on the EU's approach, something as neutral as weather could constitute "personal data" when combined with other data collected from CCTVs, WiFi tracking sensors, audio recordings, authentication systems and so on.[33] If this is indeed the case, that will mean that almost every piece of data or information would potentially trigger the application of data protection laws, something which will subsequently make it very difficult and virtually impossible for data controllers to comply with. The implication of such a scenario would be a decrease in the overall effectiveness of data protection laws as companies and organisations, particularly those with limited technological resources, would struggle to identify where to prioritize their data protection efforts. Considering this, would it not be more prudent to set the scope of "personal data" narrowly such that it is more reasonable and manageable to comply with data protection obligations rather than to pursue overly ambitious ones that are unattainable in practice?

Another problem with the expansive approach is that a single set of personal data can very easily be the personal data of numerous people at the same time, all of whom may wish to exercise different right or interests over said set of data. Take, for example, a scenario where a

[31] Sweeney, *Simple Demographics Often Identify People Uniquely* (Carnegie Mellon University, Data Privacy Working Paper 3, 2000)
[32] Purtova, '*The law of everything. Broad concept of personal data and future of EU data protection law*' (2018) 10(1) Law, Innovation and Technology 40-81
[33] Ibid

TV news reporter is being filmed and recorded on a public street and, as a result, the video produced captures the faces of many passersby. It's most likely that with an expansive approach, the video will constitute the personal data of all of the individuals captured (since information can come in the form of pictures/videos, the video relates to those passersby, and the passersby can be identified from their faces), all of whom will thus have rights over said video. This may subsequently give rise to legal complexities when, for example, one such passerby exercises his personal data rights over the video, which interferes with the privacy of other passersby and not to mention, the intellectual property rights of the TV news producer. This point is also apparent from the case of *Nowark*, where the examiner's comments were found to be both the personal data of the examiner, as well as that of the candidate.

Furthermore, interpreting "personal data" in an all-encompassing way will also lead to the challenge of distinguishing and treating different types of data appropriately. For example, it would be difficult to draw a distinction between information that unquestionably represents personal data (e.g. our date of birth, telephone number, written work/essays), and information that may not obviously reveal personal aspects but still falls within an expansive interpretation of "personal data" (e.g. a university timetable). Treating both types of information equally as personal data could be problematic, as something like our written work/essay would reveal much more about us, including our thoughts, views, and opinions, while a university timetable is primarily a schedule of classes and their respective timing. If "personal data" is interpreted so broadly, it may subsequently be necessary to generate further rules to factor in the varying degrees of privacy implications associated with different types of data to ensure a more effective approach to data protection.

Conclusion

In the face of increasing data protection challenges, it is crucial for legislators, policymakers, and judicial authorities to provide clear definitions and interpretations of "personal data" as it directly affects the material scope and application of data protection laws. The key definitional elements of "personal data" are inherently flexible and can be interpreted in various ways. While broad data protection laws are necessary in the current data-driven era, interpreting "personal data" too broadly may result in the concept encompassing virtually everything, something which might dilute and undermine the purpose of having data protection laws in the first place. Such an expansive approach may also make it necessary to create further rules to determine the extent of protection offered to different classes or types of personal data. While it may be easier to argue for an expansive approach and treat all information as "personal data" before teasing out which information should be protected to a larger extent than others, it is submitted that a narrower approach, such as that adopted by the English courts, should be preferred. A narrower approach which does not treat insignificant matters like the weather as "personal data" can ensure that data protection obligations as a whole are proportionate to the privacy risks associated with the information protected. It will also enable for a more efficient allocation of resources for companies and organisations as they would only need to focus their data protection efforts on information that are clearly understood as "personal data".