

ANONYMITY IN EU HEALTH LAW: *NOT* AN ALTERNATIVE TO INFORMATION GOVERNANCE

Summary

Data sharing has long been a cornerstone of healthcare and research, and is only due to become more important with the rise of Big Data analytics and advanced therapies. Cell therapies, for example, rely not only on donated cells but also essentially on donated information to make them traceable. Despite the associated importance of concepts such as ‘donor anonymity,’ the concept of anonymisation remains contentious. The Article 29 Working Party’s 2014 guidance on ‘Anonymisation Techniques’ has perhaps helped encourage a perception that anonymity is the result of data modification ‘techniques,’ rather than a broader process involving management of information and context.

In light of this enduring ambiguity, this paper advocates a ‘relative’ understanding of anonymity, and supports this interpretation with reference not only to the General Data Protection Regulation, but also to EU health-related legislation which also alludes to the concept. Anonymity, I suggest, should be understood not as a ‘technique’ which removes the need for information governance, but rather as a legal standard of reasonable risk-management, which can only be satisfied by effective data protection. As such, anonymity can be not so much an *alternative* to data protection as its mirror, requiring similar safeguards to maintain privacy and confidentiality.

Keywords: Anonymity, Cell Donors, Confidentiality, Data Sharing, Information Sharing,
Medical Research

I. INTRODUCTION

Anonymised data are often recognised as an integral resource for biomedical research,¹ but the concept of anonymisation is a source of enduring controversy. Anonymisation has been viewed as an avoidance of regulation,² enabling otherwise unlawful or unethical secondary uses of data,³ and allowing data to be freely used, shared, and sold.⁴ The perception that anonymisation is potentially harmful, or fundamentally unworkable, casts doubt on its utility within research,⁵ and indeed within areas such as therapeutic cell, tissue and organ donation which currently require donor anonymity as a default.

In contrast, this paper uses the words ‘anonymity’ and ‘anonymisation’ to refer to the successful maintenance of anonymity to the relevant legal standard. This standard comes from Recital 26 General Data Protection Regulation⁶ (‘GDPR’), which refers to data ‘*rendered anonymous*,’ meaning individuals are no longer identifiable by any means ‘*reasonably likely*’ to be used. It sets a test as to when data can be considered anonymous and out of scope of the GDPR. Meeting this test is better understood as a form of regulatory compliance in itself, as opposed to entry into a Wild West of Big Data flows.

¹ For example, *R. v Department of Health Ex p. Source Informatics Ltd* (No.1) [2001] Q.B. 424 (CA) 23

² M Bayern, ‘DeepMind, NHS use anonymized patient data in AI to avoid regulatory hurdles’ <<https://www.techrepublic.com/article/deepmind-nhs-use-anonymized-patient-data-in-ai-to-avoid-regulatory-hurdles/>> accessed 26 February 2020

³ J Andrew and M Baker, ‘The General Data Protection Regulation in the Age of Surveillance Capitalism’ (2019) *Journal of Business*

⁴ L Rocher, J M. Hendrickx and Y de Montjoye, ‘Estimating the success of re-identifications in incomplete datasets using generative models’ (2019) 10 *Nature Communications* 3069

⁵ B Clarke, ‘Researchers: Anonymized data does little to protect user privacy’ available from: <<https://thenextweb.com/insider/2019/07/30/anonymized-data-does-little-to-protect-privacy/>> accessed 30 July 2019

⁶ Regulation 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/ (General Data Protection Regulation) [2016] OJ L119/1, which will be cited as ‘the GDPR’

The original contribution of this paper is thus to reflect on ‘anonymity’ with reference to developments within data protection law, but also to sources of EU healthcare law governing tissue, cells and organs which also use such terminology but are less studied in connection with anonymisation. These pieces of legislation variously allude to the principle of ‘donor anonymity,’ even though it is obviously not possible for the donor’s identity to be unknown to all actors involved in the donation. For this to be feasible, I argue that a more relative interpretation of anonymity is required. The benefit of this relative interpretation is to remove focus away from eliminating identifiability *for the discloser* by modifying the data through an ‘anonymisation technique,’ and instead focus on the capacity of a specific recipient to identify individuals. This draws on existing work on the contextuality, or ‘functionality,’ of anonymisation by Elliot, Mackey, El Emam and others, identifying legal authority in support of this account, as well as highlighting why the concept is of continued relevance, even at a time when high-profile figures suggest the distinction between personal and non-personal data is obsolete.⁷

By way of background, section II begins with the Article 29 Working Party’s 2014 guidance on Anonymisation Techniques, and in particular its tension with the realities of biomedical research. Section III then considers the subsequent developments since this guidance was published, such as the transition to the GDPR, and the judgment of the Court of the European Justice of the European Union in *Breyer*.⁸ The way in which the scope of personal data was drawn in this case is, I suggest, useful when contemplating ‘donor anonymity’, which cannot be absolute anonymity. This is explored in section IV. Section V touches on the forthcoming Clinical Trials Regulation (‘CTR’), and the impact this may have on data sharing.

⁷ F Niker, ‘An Interview with Baroness Onora O’Neill (Beyond the Ivory Tower series)’ (6 January 2020) <<http://justice-everywhere.org/governance/an-interview-with-baroness-onora-oneill-beyond-the-ivory-tower-series/>> accessed 26 February 2020

⁸ Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] ECLI: EU: C: 2016:779

Finally, section VI considers the extent to which a clarified understanding of anonymisation addresses criticisms levelled at anonymity within biomedical Big Data. A more relative, organisation-specific, account of anonymity could actually help to counter the suggestion that anonymisation is an ‘opt-out’ from regulation, as it means data subjects can still exert their GDPR rights against the original provider, but can be assured to a reasonable standard that they will not be identified by third parties. However, it is accepted that this does not resolve all potential concerns, and anonymisation may in fact mirror the individually-focused weaknesses of data protection law, as well as its strengths.

Ultimately, I argue that anonymity remains a useful and important concept in EU healthcare law, not least because it is unique in indicating a state in which people will not be identified by unauthorised third parties, and this is still a distinction which could matter to patients, donors⁹ and trial participants. However, in order to be of continuing utility it must be thought of in relative terms, and in a way which takes into account information governance as a means of managing context. In healthcare research, where it is possible to regulate behaviour between collaborators through governance measures such as policies and contracts, information of greater research utility can be shared by addressing the lawful behaviour of those given access to the minimised data, and removing reasonable means of identification from them.¹⁰ Biomedical research is therefore an area which may particularly benefit from revision of the existing guidance, and a move to a more relative approach.

⁹ For example, F Mahieu, W Decleer, K Osmanagaoglu et al, ‘Anonymous sperm donors’ attitude towards donation and the release of identifying information’ (2019) 36 J Assist Reprod Genet 10; G Cohen, T Coan, M Ottey et al, ‘Sperm donor anonymity and compensation: an experiment with American sperm donors’ (2016) 3 J Law Biosci 3

¹⁰ For more detail on the balance between utility and identification risk, see M Elliot, E Mackey, K O’Hara and C Tudor, ‘The Anonymisation Decision-Making Framework’ <<https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>> accessed 26 July 2019

Terminology

For clarity, the term ‘relative’ is used in this paper to reflect the usage by the Court of Justice of the European Union to differentiate between different interpretations of the scope of personal data.¹¹ In broad terms, the objective approach sees data as personal if ‘anyone’ could identify them, whereas under the relative approach it is only the actual (be) holder of the information, and anyone they could reasonably approach, who needs to be taken into account.

I have suggested that anonymity is unique as a regulatory standard, as it assures individuals that they will not be identifiable by any means reasonably likely to be used. This is why this paper focuses on anonymity, as opposed to ‘pseudonymity.’ Under the GDPR, pseudonymisation is defined as a process of data minimisation which nevertheless does not prevent data from being personal.¹² This means that, by definition, people within pseudonymised data are still reasonably likely to be identified, and so talking of ‘donor pseudonymity’ or ‘participant pseudonymity’ would not have the same meaning. Following the current definition, only if additional steps are taken to ensure third parties are not reasonably likely to identify individuals can data be said to be anonymous for these people, as opposed to merely pseudonymised.¹³ It is a key argument of this paper that these additional steps lie in information governance, as opposed to deletion or aggregation of original data. This is considered in the next section.

II. OPINION 05/2014 ON ANONYMISATION TECHNIQUES

It has already been argued that anonymity in research cannot be objective (or ‘idealised’ as Saunders and colleagues have characterised it¹⁴) as there will always be some parties (such as

¹¹ Breyer, note 8

¹² GDPR Article 4(5)

¹³ M Mourby, E Mackey, M Elliot et al, ‘Are ‘pseudonymised’ data always personal data? Implications of the GDPR for administrative data research in the UK’ (2018) 34 CLSR 2

¹⁴ B Saunders, J Kitzinger & C Kitzinger, ‘Anonymising interview data: challenges and compromise in practice’ 15 Qualitative Research 5

the primary researchers) who will be aware of the participants' identities. Nonetheless, it may well be important for interview participants not to be identified by people outside of this confidential relationship.¹⁵ This section explores how this kind of relative anonymity within research is difficult to achieve in accordance with the Article 29 Working Party guidance on Anonymisation Techniques.

The primary legal authority on EU anonymity is Recital 26 GDPR, which provides as follows:

' [...] To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. [...]. '

This establishes a contextual question of fact as to whether data are personal or anonymous, and the GDPR does not specify any set of practices that might be used to remove data from scope.

Nonetheless, there can be a tendency to view anonymisation as a set of practices—rather than a regulatory standard—and when these practices fail, this can lead some to conclude anonymisation itself has failed,¹⁶ rather than that the term has been misapplied to a situation where the reasonable risk of identification has not been excluded, and the test has not been

¹⁵ Ibid

¹⁶ P Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCLA Law Review 1701

met.¹⁷ It is understandable, however, that the term is used in this way (whether anonymisation has actually been achieved or not) when influential EU-level guidance refers (slightly misleadingly) to ‘Anonymisation Techniques.’¹⁸ It is tempting, in the circumstances, to identify anonymisation with the techniques themselves, whether they in fact produce anonymity or not.

The Article 29 Working Party (‘A29WP’) was a body made up of representatives of supervisory authorities from each EU member state, and adopted its Opinion 05/2014 on Anonymisation Techniques on 10 April 2014. The A29WP has since been replaced by the similarly constituted European Data Protection Board (EDPB), which formally endorsed some of the A29WP’s guidance in 2018,¹⁹ but not Opinion 05/2014 itself. While 05/2014 has not been actively disavowed, and there is no evidence at the time of writing that the EDPB intends to issue any new guidance on anonymisation,²⁰ its status is less certain than if it had been formally endorsed, and is further complicated by intervening developments such as the case of *Breyer* and the transition to the GDPR.

This guidance has still been highly influential (see subsection B), despite the problematic nature of some of its requirements. For the purposes of this paper, I will focus on two issues in particular:

- 1) Its neglect of information governance;
- 2) Its suggestion that anonymity requires deletion or aggregation of original data.

¹⁷ Elliot and colleagues also argue that alleged failures of anonymisation are failures of practice, and not of law: M Elliot, K O’Hara, C Raab et al, ‘Functional anonymisation: Personal data and the data environment’ (2018) 34 CLSR 2

¹⁸ Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques WP216 (Brussels, 10 April 2014)

¹⁹ European Data Protection Board, Endorsement 1/2018

<https://edpb.europa.eu/sites/edpb/files/files/news/endorsement_of_wp29_documents_en_0.pdf> accessed 27 February 2020

²⁰ L Taranto and P Garcia, ‘Medical Research Council Advises on How to Anonymise Information for Research Purposes’ 16th October 2019, available from: <<https://www.hl dataprotection.com/2019/10/articles/international-eu-privacy/medical-research-council-advises-on-how-to-anonymise-information-for-research-purposes/>> accessed 13 November 2019

A Information Governance

As regards the first issue, the word ‘governance’ is used broadly to include any measure which helps to shape the way in which information is used. While the term ‘regulation’ has equally been used to describe the process of altering others’ behaviour by both state and non-state actors,²¹ the term ‘governance’ is preferred here as ‘information governance’ is a commonly recognised term encompassing the voluntary (and necessitated) actions taken at organisational level to control how information is used in that particular setting. The word ‘governance’ has also been used in this broader sense, encompassing leadership, cultures and relationships, within the context of biobanking.²²

In contexts such as health and life sciences research, where the relationship between data sharing parties is likely to involve a Data Access Agreement,²³ contracts would be a key example of a governance measure to help achieve, and maintain, anonymisation, by shaping and regulating how data are used. As Stalla-Bourdillon and Knight argue, when a ‘dynamic’ and environmental approach is taken to anonymisation:

‘recipients of anonymized data, although they are not data controllers when they receive the dataset, have to behave responsibly and comply with any licensing obligations imposed by the original data controllers of the raw personal data. Specifically, the former must abide by any licensing limitations upon the purpose and the means of the processing of the data in its disclosed post-anonymization process form to remain outside the scope of data protection laws. At the same time, the characterization of anonymized data should also be dependent

²¹ J Black, ‘Regulatory Conversations’ (2002) 29 Journal of Law and Society 1

²² J Kaye, SM Gibbons, C Heeney et al, Governing Biobanks – Understanding the Interplay between Law and Practice (Hart: Oxford, 1st edn, 2012)

²³ See, for example, the standard Data Access Agreement through which information is made available via the European Genome-Phenome Archive, available from: https://www.ebi.ac.uk/ega/submission/data_access_committee accessed 13 November 2019

*upon an ongoing monitoring on the part of the initial data controller of the data environment of the dataset that has undergone anonymization.*²⁴

An important way to minimise risk of re-identification of data subjects is thus a contract between data provider and recipient creating legally binding obligations such as a promise not to attempt re-identification, to maintain adequate data security, only to use data for specified purposes, access to be granted only for authorised personnel, records of processing to be kept, and specified procedures to be followed in the case of accidental re-identification. This could in effect provide a compromise by providing legal protection for data subjects, without exposing them to a reasonable likelihood of identification (which, in effect, data protection law requires for its scope).

By contrast, the A29WP guidance is implicitly dismissive of the role contracts can play in enforcing anonymisation. The only mention made comes on p.29:

‘State-of-the-art encryption can ensure that data is protected to a higher degree, i.e. it is unintelligible for entities that ignore the decryption key, but it does not necessarily result in anonymisation. For as long as the key or the original data are available (even in the case of a trusted third party, contractually bound to provide secure key escrow service), the possibility to identify a data subject is not eliminated.’

This does not explore whether, even if the trusted third party was contractually bound not to reveal the key or original data to anyone, or anyone other than an authorised recipient, this key would nonetheless be a means reasonably likely to be used for all other parties. The possibility of identification must apparently be eliminated, not just for the party receiving ‘anonymised’ information, but for *everyone* for the information to count as anonymised under this guidance. Therefore, even if it would be unlawful, and contrary to contract, for others to

²⁴ S Stalla-Bourdillon and A Knight, ‘Anonymous Data v. Personal Data – False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data’ (2016) 34 Wis. Int'l L.J. 284

obtain assistance in identifying individuals, this is presumably still a ‘means reasonably likely to be used’ in the eyes of the drafters.

Otherwise, the guidance is largely silent as to the role that due diligence, track-record, contracts, policies, training, line-management, record-keeping, accreditation, Codes of Practice and audits can play in establishing and maintaining minimal risk of identification, meaning it often reads as ‘data-centric,’ or at least ‘provider-centric’ rather than accommodating the perspective of the information recipient.²⁵ The A29WP do advise against ‘release and forget’ approaches, and recommend monitoring and control of risk.²⁶ However, by focusing the Opinion on techniques such as noise addition, randomisation, k-anonymity and pseudonymisation, they do not provide any guidance as to how this monitoring should take place, and indeed how this monitoring is in itself a safeguard against reasonable means of identification, and an important supplement to any data modifying technique. The emphasis, it seems, is not to establish trustworthy governance, but to ‘eliminate’ any technical possibility of identification prior to disclosure. It is therefore unsurprising that de-identification of datasets before sharing them has come to be seen as the paradigm for research,²⁷ rather than anonymisation being understood as an ongoing process of environmental regulation²⁸ which must be maintained after data have been transferred.²⁹

An alternative account of anonymity would see governance of anonymous data as mirroring GDPR compliance in its requirements: accountability, records of processing, purpose limitation, breach procedures and data minimisation would remain core aspects of risk-

²⁵ M Elliot and A Dale, ‘Scenarios of attack: the data intruder’s perspective on statistical disclosure risk’ (1999) 14 Netherlands Official Statistics 6

²⁶ Note 18, p.24

²⁷ Rocher et al, note 4

²⁸ E Mackey and M Elliot, ‘Understanding the Data Environment’ (2013) XRDS <[10.1145/2508973](https://doi.org/10.1145/2508973)> accessed 16 December 2019

²⁹ For more detail as to how, see M Elliot, E Mackey, K O’Hara and C Tudor, ‘The Anonymisation Decision-Making Framework’ <<https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf>> accessed 26 July 2019

management for anonymised data post-disclosure, to ensure no reasonable means of identification re-emerges. It would be helpful if these dimensions of anonymisation were more widely recognised. Controllers entering into a Data Processing Agreement, for example, benefit from helpful guidance, and indeed a checklist within Article 28 GDPR, as to what the contract should contain. The UK Medical Research Council does provide some guidance as to what an Agreement to support anonymisation could include, such as a prohibition on identification attempts and contingency terms in the case of accidental identification.³⁰ For the sake of data sharing across the EU, this could be developed further at EDPB level, and involve considerations such as:

- Has the discloser exchanged information with the recipient before?
- Is there any reason to doubt their track record of reliable use of data?
- Has the recipient provided evidence of appropriate training or policies as to safe data handling within their organisation?
- Is a data access agreement in place? Does said agreement:
 - Prohibit re-identification
 - Prevent all but a specified number of uses
 - Prohibit linkage with other information, or onward sharing of the data
 - Place a time-limit on the retention of the data
 - Limit the number of people who can access the data
 - Limit the purposes for which they can use the data
 - Require the recipient to keep records of their use of the data
 - Allow the discloser to audit the use of the data
 - Require the recipient to delete the data at the discloser's request

³⁰ Medical Research Council, Regulatory Support Centre, 'Identifiability, anonymisation and pseudonymisation: Guidance note 5' available from: <https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-5-identifiability-anonymisation-and-pseudonymisation/> accessed 13 November 2019

- Make provision for the procedure in the event of accidental re-identification?
- To the extent that preparing the data for disclosure requires internal processing of personal data—has the discloser been transparent with their data subjects about this processing, and its ultimate purpose?

The above questions are of the kind discussed at an Anonymisation workshop hosted by the Medical Research Council in 2019, although this is by no means exhaustive.³¹ Likewise, while contracts are by no means the sole mechanism through which identifying behaviours could be managed, they provide an example of the way in which information governance can control context, and thus the means reasonably likely to be used by a recipient. It would be valuable if any new guidance from the EDPB could acknowledge this.

B Deletion/ Aggregation of Original Data

The suggestion that the original information should be deleted or aggregated is another respect in which the A29WP set an unhelpful precedent, a key quotation from the guidance being:

‘Secondly, “the means likely reasonably to be used to determine whether a person is identifiable” are those to be used “by the controller or by any other person”. Thus, it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data. Only if the data controller would aggregate the data to a level where the individual events are no longer identifiable, the resulting dataset can be qualified as anonymous.’³²

³¹ ‘Safe Sharing of Research Data: The role of legal agreements when anonymising’, 25th April 2019, IET London. Any errors or misapprehensions are the author’s own.

³² Note 18, p.9

This passage makes it clear that the A29WP considered data to be personal if they could be identified by ‘any’ person, including the original controller. This is important as the CJEU implicitly approved, in 2016,³³ the submission that personal data need only be identifiable by those the controller could ‘reasonably approach’ (see Section III). As such, it is questionable whether this requirement is still tenable. If a data recipient could not reasonably approach the provider of the information for assistance in identifying individuals (for example, if it was illegal or contrary to contract) it seems unnecessary and impracticable for the latter to delete their information. This removes from the scope of anonymous data sharing many holders of medical or clinical trial records who naturally cannot delete original data.

Even before the *Breyer* judgment, the deletion requirement was criticised for its impracticality, with El Emam and Alvarez arguing forcefully in 2014:

‘The implications of this interpretation are quite severe because some projects and programs will still need the original data to conduct their business. For example, consider a hospital that wished to provide anonymized data for research. The hospital needs to retain the original data because that original data are required to treat the patients. To destroy or aggregate the original data would not make any sense.’³⁴

This is as true now as it was in 2014, when the A29WP guidance was first adopted. It has meant that, for the last five years, anyone following this guidance would have to inform data subjects that their identifiable information will be shared with third parties, if it is not possible to delete it at source. Yet this guidance remains influential, with the European Medicines Agency (‘EMA’) advising in the context of clinical trial data:

³³ *Breyer*, note 8

³⁴ K El Emam, C Álvarez, ‘A critical appraisal of the Article 29 Working Party Opinion 05/2014 on data anonymization techniques’ (2015) 5 International Data Privacy Law 1

‘Pseudonymisation reduces the linkability of a dataset with the original identity of a data subject but when used alone will not result in an anonymous dataset, therefore data protection rules still apply. It is, therefore, important to clarify that pseudonymisation is not an anonymisation method but a useful security measure. Consequently, additional measures should be considered in order to render the dataset anonymised, including removing and generalising attributes or deleting the original data or at least bringing them to a highly aggregated level.’³⁵

This passage closely echoes the A29WP guidance cited above in proposing deletion or aggregation of original data to achieve anonymity, as opposed to controlling the context into which information is disclosed.³⁶ This proposal is particularly problematic within guidance focused on the *publication* of data, as reducing the prospect of identification from the original data does not affect the ability of others to identify individuals by other means.³⁷ Also, as the European Commission has pointed out in relation to serious adverse reactions, it is unrealistic to expect original clinical trial data to be capable of deletion, and so El Emam’s objection has thus been tacitly acknowledged by the EC (albeit in the context of deletion where consent has been withdrawn).³⁸ This highlights another way in which the A29WP guidance is not the most helpful benchmark for research data. In a controlled access context, where it would be possible to ensure, to a reasonable standard, that the third party cannot or will not identify individuals from the data, it seems unnecessary and undermining of confidence not to be able

³⁵ European Medicines Agency, ‘External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use,’ (2018) version 1.4 <https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-3.pdf> accessed 25 February 2020, p.41.

³⁶ Contractual control, and other means of managing context, are acknowledged on the preceding page of the guidance (p.40) so it is unclear whether the EMA considers the deletion requirement to be engaged in all circumstances.

³⁷ See Rocher et al, note 4

³⁸ European Commission Directorate-General for Health and Food Safety, ‘Question and Answers on the interplay between the Clinical Trials Regulation and the General Data Protection Regulation, page 7, available at: <https://ec.europa.eu/health/sites/health/files/files/documents/qa_clinicaltrials_gdpr_en.pdf> accessed 29 January 2020

to tell people that their shared information will be anonymous, especially as there is evidence to suggest anonymity is an important condition for public approval of health data sharing.³⁹

Neglect of information governance and an emphasis on deletion are only two aspects of the A29WP guidance. However, they are elements which help to explain the perception that anonymisation involves the modification and subsequent neglect of data. As the above citation from the EMA demonstrates, they are also elements which continue to prove influential within EU healthcare practice. While this is problematic in its own right, developments since 2014 also challenge this understanding of anonymity. These developments are explored in the next section.

III BREYER AND THE GDPR

In assessing whether the A29WP guidance is still a useful benchmark, a key question is therefore whether anonymisation needs to render data incapable of identification by *anyone*? If the answer to this question is ‘no’, the deletion requirement is essentially defunct, information governance within a particular context becomes far more relevant, and it is far easier to think of anonymity in relative terms. This in turn assists with the concept of ‘donor anonymity,’ so it is an important question to consider first.

An initial, if seemingly minor, point to make is that Recital 26 GDPR uses language very similar to that of the former Data Protection Directive 95/46 EC (‘the Directive’),⁴⁰ but not identical. Recital 26 of the Directive provided:

‘Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should

³⁹ M Aitken, J de St. Jorre, C Pagliari et al, ‘Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies’ (2016) 17 BMC Medical Ethics 73

⁴⁰ 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, OJ 1995 L 281/31 (‘the Directive’)

*be taken of all the means likely reasonably to be used either by the controller **or by any other person** to identify the said 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...*’ (emphasis added).

This is the language on which the 2014 A29WP Anonymisation guidance was based. The similar, but subtly different, text of Recital 26 GDPR reads:

*‘The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller **or by another person** to identify the natural person directly or indirectly’* (emphasis added).

The difference between ‘any other’ person and ‘another’ person may appear trivial, but it goes to the heart of the difference between an objective and a relative understanding of anonymity. It was, in essence, the question at the heart of the *Breyer* judgment: are data personal if they can be identified by ‘any’ person, or only if they can be identified by the controller and those they are reasonable likely to approach for assistance (‘another’ person)?

The case related to whether the German government had at their disposal means to identify individuals from IP addresses. The case was referred, and decided, under the Directive, and so the old language of ‘any other person’ in Recital 26 still applied. Even without the GDPR’s clarifying switch to ‘another’ person, the seemingly objective phrase ‘any other’ was not taken literally. If an entirely objective view had been taken, the theoretical ability of the

Internet Service Provider (ISP) to identify IP addresses would have been sufficient, and only if they deleted their client records could the government data have been out of scope.

Instead, Advocate General Campos Sanchez-Bordona argued that apparently absolute statements within Recital 26 of the Directive should be understood in light of the ‘means reasonably likely to be used’ benchmark, observing:

‘The expression ‘means likely reasonably to be used ... by any other person’ ... could give rise to an interpretation according to which... it would be sufficient that any third party might obtain additional data [...].

65. *That overly strict interpretation would lead, in practice, to the classification as personal data of all kinds of information, no matter how insufficient it is in itself to facilitate the identification of a user. [...]*

68. *Just as recital 26 refers not to any means which may be used by the controller... but only to those that it is likely ‘reasonably’ to use, the legislature must also be understood as referring to ‘third parties’ who, also in a reasonable manner, may be approached by a **controller** seeking to obtain additional data for the purpose of identification. This will not occur when contact with those third parties is, in fact, very costly in human and economic terms, or practically impossible or prohibited by law⁴¹ (emphasis added).*

Paragraph 68 of the Advocate’s Opinion, above, was cited approvingly by the Court of Justice of the European Union in their subsequent judgment.⁴² The IP addresses were still found to be personal, but only because the German government had lawful means to acquire information from the ISP’s to identify them. The court apparently heeded the Advocate General’s warning that it is impossible to say with certainty that there is no theoretical third

⁴¹ Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland [2016] ECLI: EU: C: 2016:779, Opinion of AG Sanchez-Bordona, paras 64-68

⁴² Note 8, para 46

party capable of revealing the identity of data subjects. Strict observance of the phrase ‘*any other person*’ could therefore make a mockery of the ‘means reasonably likely to be used’ qualification. Under Sanchez-Bordona’s argument, therefore, data are personal if they can be identified:

- 1) By the ‘controller’ of the information (albeit not necessarily ‘data controller’ if the data are not personal⁴³) and/or;
- 2) By, or with the assistance of, those parties whom said ‘controller’ could ‘reasonably’ approach for aid in identifying individuals.

This pool of parties the controller could reasonably approach for assistance in identification is limited. Cost, both human and economic, should be taken into account—human cost potentially encompassing not only time but also reputational risk. Practical impossibility or prohibition by law are also relevant factors. In instances where the controller is prohibited by contract (or by law⁴⁴) from attempting re-identification, it seems unlikely that they could successfully approach the original data provider and request assistance in identifying data subjects.

Therefore, where such an approach would constitute breach of contract, or even an attempted criminal offence, it seems much less likely that these parties should be legitimately placed in the camp of third parties who can help identify individuals. As Purtova notes, while legal prohibition cannot mean individuals are definitely not identifiable,⁴⁵ as the risk of illegal acts should not be discounted, legal prohibition is a factor which may make re-identification *less*

⁴³ Under Article 4(7) GDPR, a data controller is the entity which determines the purposes and means of processing personal data. When an entity only controls anonymous data it is strictly speaking not a data controller according to this definition, but the term ‘controller’ is still a convenient term to denote their use of the information.

⁴⁴ For example, it is a criminal offence under section 117 of the UK’s Data Protection Act 2018 to re-identify de-identified data without the consent of the data controller.

⁴⁵ Indeed, in light of the above-referenced UK provision, this would mean that all de-identified data in the UK would be automatically considered anonymous for everyone except the data controller and those they authorise to re-identify individuals, thus eliminating the need for security standards and information governance.

reasonably likely.⁴⁶ However, when combined with adequate governance measures designed to prevent behaviour which could lead to re-identification (contracts, policies, supervision, professional ethical codes of conduct as well as legal prohibition), it may be sufficient to say that a controller does not have access to means reasonably likely to be used to identify individuals, either on their own or through reasonable approaches to third parties.

It therefore follows that, even if data could be theoretically identifiable for one party, this does not render them personal for another if they cannot identify people, or reasonably approach others for assistance in identification. This conclusion is hugely helpful for the concept of donor anonymity, where donors and recipients are ‘anonymous’ vis-à-vis each other, but not to their respective healthcare providers. This calls into question any deletion requirement for anonymisation, and opens up considerations of context in assessing risk of identification. The utility of a more relative account of anonymity is explored in the next section.

IV DONOR ANONYMITY

This section considers the therapeutic use of donated cells, tissues and organs. Here, a relative account of anonymity is coherent with the principle of donor anonymity as posited in the legislation. However, unlike healthcare research, the therapeutic relationship with the patient is unlikely to be regulated by a contract in which a patient’s use of information can be circumscribed. As such, anonymity through information governance in this context appears to entail the non-disclosure of information as a default, and severely restricted sharing as an exception.

⁴⁶ N Purtova, ‘The law of everything. Broad concept of personal data and future of EU data protection law’ (2018) 10 Law, Innovation and Technology 40

The Tissues and Cells Directive⁴⁷ provides a framework for the national law of EU members as regards the donation of tissues and cells. It is supplemented by the directly effective Advanced Therapy Medicinal Product Regulation,⁴⁸ which also regulates somatic cell and tissue engineered therapies, which may use donated cells. It also sits alongside the Organ Donation Directive,⁴⁹ which similarly regulates information relating to organ donors. All three of these pieces of legislation refer, in one way or another, to the principle of donor anonymity.⁵⁰

Article 14(1) of the Tissues and Cells Directive, for example, is headed ‘Data protection and confidentiality.’ It appears to allude to anonymity in the same sense as the term is used under data protection law. It deploys the same terminology of ‘rendered anonymous’ as used in the former Data Protection Directive, and is now used in the GDPR:

*‘Member States shall take all necessary measures to ensure that all data, including genetic information, collated within the scope of this Directive **and to which third parties have access, have been rendered anonymous** so that neither donors nor recipients remain identifiable’* (emphasis added).

It seems clear that the term ‘rendered anonymous’ is intended to have the same meaning as in data protection legislation, as it comes in an Article headed ‘Data protection’ which uses the same language. In only requiring data ‘*to which third parties have access*’ to be rendered anonymous, it also seems to draw on a relative interpretation of anonymity, in which data can be ‘rendered anonymous’ for a third party but remain personal data for the controller. As the Article goes on to say:

⁴⁷ Council Directive 2004/23/EC of 31 March 2004 on setting standards of quality and safety for the donation, procurement, testing, processing, preservation, storage and distribution of human tissues and cells [2004] OJ L 102/48

⁴⁸ Council Regulation (EC) 1394/2007 of 13 November 2007 on advanced therapy medicinal products and amending Directive 2001/83/EC and Regulation (EC) No 726/2004 [2007] OJ L324/121

⁴⁹ Council Directive (EU) 2010/45 of 7 July 2010 on standards of quality and safety of human organs intended for transplantation [2010] OJ L 207/14

⁵⁰ Note 47 Recitals 18 and 29, note 48 Recitals 15 and 19, note 49 Recital 22

‘2. For that purpose, they shall ensure that:

(a) data security measures are in place, as well as safeguards against any unauthorised data additions, deletions or modifications to donor files or deferral records, and transfer of information;

(b) procedures are in place to resolve data discrepancies; and

*(c) no unauthorised disclosure of information occurs, **whilst guaranteeing the traceability of donations.***’

3. Member States shall take all necessary measures to ensure that the identity of the recipient(s) is not disclosed to the donor or his family and vice versa, without prejudice to legislation in force in Member States on the conditions for disclosure, notably in the case of gametes donation’ (emphasis added).

2(c) is particularly interesting, as it makes it clear that when information is provided to third parties, the act of rendering data anonymous cannot include the deletion of the original data, as the traceability of the donor and recipient must be maintained. This is clearly incompatible with the deletion/ aggregation requirement in the A29WP guidance. The guidance would require data to be rendered anonymous for the controller, even though they are authorised to hold this confidential information, and their ability to identify subjects does not necessarily equate to that of a third party. This requirement would be particularly impracticable in the context of tissue and cell donors and recipients, who must remain traceable at least by those authorised to hold their confidential information. Even though there is no question of the donor or recipient’s medical safety being compromised by deletion of their own healthcare records, the word ‘anonymous’ is nonetheless used in the Tissues and Cells Directive. This suggests that the two parties do not need to be anonymous for everyone, to be anonymous to each other.

It could be argued that, as the A29WP was specifically turning its mind to anonymisation, and the drafters of the Cells and Tissues Directive were not, the 2014 anonymisation guidance should prevail, and the Cells and Tissues Directive should be amended accordingly. Viewed from this perspective, the word ‘anonymous’ should be understood in a loose sense within the Cells and Tissues Directive, and should perhaps be replaced with the word ‘pseudonymous’ instead.

However, there are two problems with the contention. Firstly, if the term ‘anonymous’ or ‘anonymity’ were removed from the Cells and Tissues Directive, it would be difficult to ensure donors were protected to the desired standard. If the word ‘anonymous’ were replaced with ‘pseudonymised’ or ‘de-identified,’ there would be no accompanying requirement for the information disclosed to be other than personal data, i.e. capable of identifying individuals by means reasonably likely to be used. The protection offered to subjects would thus be weaker if all the Directive required was for data providers to render them less identifiable, without any thought as to whether they are still reasonably likely to be identified by the recipient. This level of identity protection is unique to the term ‘anonymous’; without this standard subjects cannot be offered a reasonable level of confidence that their identity will not be revealed to third parties.

Secondly, as observed at the beginning of this subsection, the Tissues and Cells Directive deliberately uses the language of rendering data anonymous, and was thus evidently not intended to be used in a loose sense. A more compelling interpretation is that the Tissues and Cells Directive, like the Organ Donation Directive and the ATMP Regulation, refers to donor anonymity meaning these individuals should not be identified by each other, or any other unauthorised parties, by any means reasonably likely to be used. In other words, information disclosed to a non-authorised party should not reveal their identity by means reasonably likely to be used, but they do not need to be (and cannot be) anonymous for everyone. Recital

29 of the Tissues and Cells Directive therefore appears to use the word ‘anonymity’ in a relative (or ‘subjective’) sense:

‘As a general principle, the identity of the recipient(s) should not be disclosed to the donor or his/her family and vice versa, without prejudice to legislation in force in Member States on the conditions of disclosure, which could authorise in exceptional cases, notably in the case of gametes donation, the lifting of donor anonymity.’

The meaning of ‘anonymity’ must correlate to the similar instruction in Recital 22 of the Organ Donation Directive:

‘As a general principle, the identity of the recipient(s) should not be disclosed to the donor or the donor's family or vice versa, without prejudice to legislation in force in Member States which, under specific conditions, might allow such information to be made available to donors or donors' families and organ recipients.’

Changing this text to refer to ‘donor pseudonymity’ would not prevent disclosure of information reasonably likely to identify these individuals. Requiring raw data to be deleted, or aggregated, to achieve anonymity for all parties would be clearly contrary to the requirement of traceability. This is strongly supportive of a relative account of anonymity, without which anonymity in the context of traceable donations would be impossible.

In short, the principle of donor anonymity illustrates the limited, relative sense in which the word ‘anonymous’ is used in EU legislation beyond the GDPR. It is a relatively straightforward proposition where donors and recipients are given little or no information about each other. As explored further in the next section, anonymity is more challenging when information is publicly accessible, and therefore no longer subject to information governance.

V. CLINICAL TRIALS REGULATION

The Clinical Trials Regulation⁵¹ or ‘CTR’ came into force in 2014 with the aim of simplifying and harmonising the governance of clinical trials in the European Union.⁵² A core aspect of the CTR is the creation of a central portal through which applications for authorisation of clinical trials, and subsequent mandatory information, can be sent by sponsors to the relevant body within each member state.⁵³

Information submitted through the portal will be stored in a publicly accessible database, termed ‘the EU database.’⁵⁴ This EU database is intended to include mandatorily submitted information about clinical trials, but not the personal data of trial participants. Clinical Trial participants therefore provide another case study of relative anonymity within health law, as they should be identifiable within trials but not to the world at large.

A Information submitted to the EMA

The CTR provides that annual safety reports of investigational medicines should be submitted to the EMA, and stipulates:

‘3. The annual report referred to in paragraph 1 shall only contain aggregate and anonymised data.’⁵⁵

This is striking in its use of the term ‘aggregate and anonymised’ data. In contrast to ‘donor anonymity,’ it is less clear what is meant by ‘anonymised’ in this context. It could mean that aggregate and anonymised data are different types of information, or that anonymisation requires aggregation, depending on whether the word ‘and’ is conjunctive or disjunctive.

Under a relative account of anonymity, however, it would be possible for individual-level

⁵¹ Council Regulation (EU) 536/2014 of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC [2014] OJ L 158/1 (hereafter cited as ‘CTR’)

⁵² CTR, Recital 4

⁵³ CTR, Article 80

⁵⁴ CTR, Article 81

⁵⁵ CTR, Article 43

information to be submitted to the EMA, where necessary, while still qualifying as anonymised.

While individual-level information should obviously be kept to the minimum necessary for the safety report, a sponsor providing limited individual-level information (stripped of direct identifiers) to a public authority is in a very different position to private citizens receiving donor information from a public authority. The focus in the CJEU's *Breyer* judgment on the lawful means the German government had at their disposal to obtain additional information to identify individuals was key to their finding that the dynamic IP addresses of those visiting government-controlled websites were personal data. The converse inference is that, if the German government had no lawful means of identification, the IP addresses would not have been personal data in their hands. While it should not be inferred as a more general rule that all illegal re-identification is therefore not reasonably likely,⁵⁶ there is a more compelling argument that public authorities are vested with a level of trust to behave within the law, which goes above and beyond the expectations placed on ordinary citizens and corporations. The re-identification safeguards a discloser should impose should be adjusted accordingly.

It is naturally not realistic to expect sponsors providing clinical safety reports to demand the right to audit the EMA to ensure their data subjects are not re-identified, as might be the case in medical research collaboration (see Section II). But, at the same time, courts may well judge it reasonable for a sponsor to trust the EMA not to break the law. It could therefore be reasonable to submit minimised, individual-level information to the EMA where necessary for a safety report, on the understanding that the Agency is not authorised to re-identify people from this minimised information. This trust is arguably as important a safeguard as any data modification technique. The same argument, however, does not work when

⁵⁶ Purtova, note 46

individual-level information are made available to the public at large, which could be accessed by any range of trustworthy or non-trustworthy actors.

B Information Made Publicly Accessible

The compromise which appears to emerge from the above is that it could be possible for information to be provided to the EMA on an anonymous basis. However, information made publicly accessible in the EU database is more difficult to anonymise reliably as the perspective of the potential recipients of that information is more difficult to anticipate.

The CTR appears to accommodate this, as Article 81(7) CTR stipulates that *‘No personal data of subjects shall be publicly accessible.’* By implication, although the EU database will be publicly accessible by default, where personal data of trial participants could be exposed by the release of information, that data should be excluded from the publicly accessible aspects of the register. For example, Recital 67 GDPR suggests that clinical study reports should be included in the EU database:

‘The EU database should be publicly accessible and data should be presented in an easily searchable format, with related data and documents linked together by the EU trial number and with hyperlinks, for example linking together the summary, the layperson’s summary, the protocol and the clinical study report of one clinical trial’ (emphasis added).

Examination of the Medicinal Products Directive⁵⁷ (which specifies the contents of clinical study reports) and the EMA’s own guidance⁵⁸ suggests that clinical study reports should contain individual-level information, for example including information about any participants prematurely withdrawn from a clinical trial. The EMA notes that, in respect of

⁵⁷ Council Directive 2001/83/EC of 6 November 2001 on the Community code relating to medicinal products for human use [2001] OJ L311/67, Annex 1, Module 5

⁵⁸ European Medicines Agency, ‘NOTE FOR GUIDANCE ON STRUCTURE AND CONTENT OF CLINICAL STUDY REPORTS’ <https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-3-structure-content-clinical-study-reports-step-5_en.pdf> accessed 19 July 2019

each participant discontinued after enrolment (who should be identified by a patient identifier):

‘It may also be useful to include other information, such as critical demographic data (e.g. age, sex, race), concomitant medication, and the major response variable(s) at termination.’⁵⁹

This would be problematic if this detailed information about participants were made available to the public at large, given the requirement that no participant personal data should be recorded in the public database.⁶⁰ It is unlikely to be appropriate to call such individual-level information anonymous, unless a strong argument could be made that the information is insufficiently unique for anyone – including the subject, their family or doctor – to identify them. Even this argument is susceptible to miscalculation, however, and it would be preferable that ‘anonymous’ data were only shared with a party subject to strict re-identification controls, including to a public authority who should be trusted to behave within the law, and not illegally re-identify subjects. To return to the test discussed in section III, it is no longer possible to say who the ‘controller’ is, and whom they might reasonably approach, if said controller is a theoretical ‘anyone’ within the general public, and so without information governance the relative account of anonymity no longer functions.

This supports the case for submission of any necessary individual-level information relating to clinical trial subjects to the EMA, but not publishing such information on an open, ‘anonymised’ basis. Public disclosure without governance controls, solely reliant on data modification techniques, risks perpetuating the longstanding trend of re-identification from publicly released information.⁶¹ This in turn risks fuelling criticism of anonymity by

⁵⁹ Ibid

⁶⁰ CTR, Article 81(7) and Recital 67

⁶¹ Ohm, note 16 & Rocher et al, note 4

mischaracterising it as a set of techniques, rather than a legal standard which should reasonably be met through a more holistic assessment and management of information use.

The next section addresses criticisms of anonymisation, and considers the extent to which these stem from the way it has been perceived as a potentially unsuccessful set of techniques. This ultimately culminates in a defence of the continuing utility of the standard, as an important benchmark as to when subject identities are (not) protected.

VI. CRITICISMS OF ANONYMISATION

In light of the above exploration of the term ‘anonymous’ in data protection and other EU law, anonymisation can be described as the result of practices which:

- Ensure through governance as well as data modification that information revealed to a third party does not relate to any natural person they can identify (by means reasonably likely to be used);
- As such, mean that the confidentiality of the information has been protected;
- Mean that, strictly speaking, the recipient does not need to comply with the GDPR.

However, it would be difficult to argue that reasonable means of identification have been excluded if the governance does not involve significant restrictions on the use of individual level data, which may in practice closely mirror GDPR requirements.

Correspondingly, anonymisation does *not*:

- Enable the recipient to do whatever they like with the data, as this would be contrary to any adequate controls against re-identification.
- Mean that the data controller is not processing personal data in making the disclosure—the data are still identifiable *for them* at the point of disclosure. The

disclosure should therefore still comply with the GDPR, including provisions relating to transparency.

- Resolve the ambiguity as to whether the discloser remains a data controller of ‘anonymised’ data, where they have set strict terms for its use.⁶²

In short, anonymisation is a means of preserving confidentiality by protecting individual identities to a reasonable standard. It cannot be achieved without multi-faceted attempts to limit what is done with individual-level data. As such, it does not represent a more liberal regime of data processing, but merely one in which information can be disclosed to a particular recipient without specific consent or breach of the duty of confidence.

In light of this proposed characterisation, this section briefly addresses some of the longstanding, but also enduring, criticisms of anonymisation, in order to clarify the extent to which this characterisation of anonymisation addresses these concerns.

To start with, it should be acknowledged that even a relative, governance-based approach to anonymisation will not prevent what have been characterised as group harms. Floridi, for example, has argued that it is a ‘*very dangerous fallacy to think that if we protect personal data that identify individuals, the protection of the groups will take care of itself.*’⁶³ Likewise, the protection of anonymised individual-level data does not prevent inferences drawn at group level being re-applied to the same or other individuals in a harmful way. Taylor highlights genetic data as an example of information which relates to the privacy of whole minority ethnic groups, and for which individual-level consent or identity protection offers

⁶² UK caselaw suggests this is not the case, and the discloser would only be the controller for their own internal processing, with subsequent processing by the recipient outside the scope of data protection law e.g. *Common Services Agency (CSA) v Scottish Information Commissioner* [2008] UKHL 47. This may be debatable, however, if the discloser imposes conditions on the use of what are (for them) personal data.

⁶³ L Floridi, ‘Open Data, Data Protection, and Group Privacy’ (2014) 27 *Philos. Technol.* 1

inadequate redress.⁶⁴ Inferences drawn from personal or anonymised data may be easily detached from the individual-level information from which they were derived, and formulated at a generalised level, but nonetheless re-applied to individuals when they are stratified into (for example) categories supporting decisions relating to health insurance or even prescribing practices.⁶⁵ It has therefore been argued with some justification that inferential data pose the greatest risks in terms of privacy and discrimination, but are offered the least protection under data protection law.⁶⁶

To the extent that measures to prevent re-identification mirror the protections required by the GDPR, and anonymisation mirrors data protection, anonymisation will also replicate data protection's weaknesses and limitations. There is the same focus on individual-level rights, and an inability to govern the intellectual consequences of data processing. This is why exploration of law against data-driven discrimination,⁶⁷ and discussion of more systemic oversight of Big Data in health,⁶⁸ or even Harm Mitigation Bodies to scrutinise downstream effects of data use,⁶⁹ are of increasing value as use of advanced analytics increases, and legal anonymity can only be a solution to a limited set of problems. The Anonymisation Decision-Making Framework rightly addresses this by incorporating ethical reflection as part of the anonymisation process,⁷⁰ which could encompass debate on potential group harms, but this

⁶⁴ M Taylor, *Genetic Data and the Law: A Critical Perspective on Privacy Protection* (CUP: Cambridge, 1st edition, 2012), 150-151

⁶⁵ M Ravindranath, 'How your health information is sold and turned into 'risk scores'' <https://www.politico.com/amp/story/2019/02/03/health-risk-scores-opioid-abuse-1139978?__twitter_impression=true> accessed 26 July 2019

⁶⁶ S Watcher and B Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law' (2019) *Colum. Bus. L. Rev.*

⁶⁷ Although such laws may also have significant limitations—see W Nicholson Price II and I Glenn Cohen, 'Privacy in the age of medical big data' (2019) 25 *Nature medicine* 37

⁶⁸ E Vayena and A Blasimme, 'Health Research with Big Data: Time for Systemic Oversight' (2018) *The Journal of Law, Medicine & Ethics* <<https://journals.sagepub.com/doi/10.1177/1073110518766026>> accessed 2 August 2019

⁶⁹ A McMahon, A Buyx, B Prainsack 'Big Data Needs More Collective Responsibility: The Role of Harm Mitigation in the Governance of Data Use in Medicine and Beyond' (2019) *Med Law Rev* <<https://academic.oup.com/medlaw/advance-article/doi/10.1093/medlaw/fwz016/5543530>> accessed 7 August 2019

⁷⁰ Note 29

goes above and beyond what is required to manage identifiability to the legal standard of anonymity, and cannot be guaranteed as an aspect of legal compliance.

A governance-based approach does, however, help to address the concern that anonymisation creates a Wild West of Big Data, in which artificial manipulations of data can enable any and all subsequent uses. It prevents some of the danger posed by data-centric, governance-light ‘release and forget’ models in which assessing identifiability from the perspective of the discloser creates a false sense of objectivity, as though making data less identifiable *for them* will necessarily make them non-identifiable for everyone else. Given the prominence of such modification-based approaches, it is not surprising that anonymisation has been described as a technical solution, rather than an ethical one.⁷¹ Neither is it surprising that some see the distinction between personal/ anonymous data as being determined at point of collection (or disclosure) and then ignored as identification risk fluctuates unchecked within subsequent usage.⁷² Andrew and Baker, for example, argue that the law lacks vital understanding of contemporary big data practices, and this may be attributable to an overzealous trust in technical solutions.⁷³

These are all indeed legitimate criticisms if modified, individual-level data are disclosed to one or many parties without assessment and control of re-identification risk in the new environment(s), and this process is referred to as ‘anonymisation.’ It is a central argument of this paper, however, that such an act cannot be seen as rendering data anonymous. Whether there exists a means reasonably likely to identify someone is inevitably a function of the context in which those data are situated. As stated in a UK House of Lords judgment on the release of modified medical information relating to children:

⁷¹ G Laurie and L Stevens, ‘Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom’ (2016) 43 Journal Law and Society 3

⁷² S Watcher, ‘Data protection in the age of big data’ (2019) 2 Nature Electronics 6

⁷³ J Andrew and M Baker, ‘The General Data Protection Regulation in the Age of Surveillance Capitalism’ (2019) Journal of Business

‘Whether or not the individuals are identifiable from the barnardised data is a question of fact, the answer to which may vary from situation to situation and, indeed, from individual to individual.’⁷⁴

If there is an overzealous trust in technical solutions, this is not the fault of the law but of those seeking to apply it. The law has set a reasonable standard, to achieve the explicitly limited purposes of privacy and data protection, but its implementation relies on adequate assessment of identification risk, which is a question of fact. Those who ignore context, and the extent to which such context can be adequately reviewed and controlled through information governance, do the law a disservice.⁷⁵ Technical solutions cannot be a substitute for an analysis of circumstance, and the law does not suggest they should be. To return to the Advocate General’s opinion in the *Breyer*, factors which should be taken into account in determining whether data are personal include consideration of third parties a controller may approach, human and economic cost, practical (im)possibility and prohibition by law.⁷⁶ Neither the GDPR itself nor legal authority determined under the Data Protection Directive promotes a context-free approach to anonymisation.

Relational Anonymity?

It is understandable that discussion of the re-use of health-related information raises anxiety about severing the relationship between patients and their data,⁷⁷ and anonymisation (at least in its absolute form) can be such a severance between information and the rights formerly associated with it. However, as Ballantyne has argued, a better response to this risk is not to entrench connection through ideas of individual ownership, but instead develop flexible

⁷⁴ *Common Services Agency v Scottish Information Commissioner* [2008] UKHL 47, para 87

⁷⁵ Again, see Elliot and colleagues on the distinction between failures in privacy law and practice (note 17)

⁷⁶ Note 41

⁷⁷ A Ballantyne, ‘How should we think about clinical data ownership?’ (2020) *J Med Ethics* <10.1136/medethics-2018-105340> accessed 7 February 2020

models to reconnect patients with their data,⁷⁸ with emphasis on consent, transparency and engagement.

A relative understanding of anonymity could fit within such a flexible model. If data are personal for the organisation collecting the information, but not for the third parties with whom it is shared on a controlled access basis, the collector is still subject to the GDPR and subjects can enforce their rights against this entity, without exposure to identification risk from third parties.

Furthermore, it is important to remember that the rights of privacy, confidentiality and data protection we fear severing are themselves contextual, and even relational. Taylor and Wilson have demonstrated that the touchstone for a (medical) duty of confidence is now a reasonable expectation of privacy, which (like anonymity) is determined based on all circumstances of the case.⁷⁹ They argue that careful respect for autonomy is vital for compliance with these reasonable expectations, although in practice this autonomy may be of a more collective, relational nature⁸⁰ that could also be accommodated within ‘flexible’ transparent models of data use.

A useful illustration of the shift towards relational privacy comes from the Court of Appeal of England and Wales in *R (W, X, Y and Z) v Secretary of State for Health and Secretary of State for the Home Department*:⁸¹

‘the question whether there is a reasonable expectation of privacy is a broad one which takes account of all the circumstances of the case. We do not see how overseas visitors who, before they are treated in an NHS hospital, are made aware of the fact that, if they incur charges in

⁷⁸ Ibid

⁷⁹ MJ Taylor and J Wilson, ‘Reasonable Expectations of Privacy and Disclosure of Health Data’ (2019) 27 Medical Law Review 3

⁸⁰ E Dove, SE Kelly, F Lucivero et al, ‘Beyond Individualism: Is there a place for relational autonomy in clinical practice and research?’ (2017) 12 Clinical Ethics 3

⁸¹ *R (on the application of W, X, Y and Z) v Secretary of State for Health and Secretary of State for the Home Department, the British Medical Association* [2015] EWCA Civ 1034, [44]

excess of £1,000 and do not pay them within 3 months, the Information may be passed to the Secretary of State for onward transmission to the Home Office for the stated immigration purpose can have any, still less any reasonable, expectation that the Information will not be transmitted in precisely that way. They will, however, have a reasonable expectation of privacy in relation to the Information vis-à-vis anyone else.'

There is an alignment between this characterisation of privacy as arising vis-à-vis some people, but not others, and the idea of anonymity as existing vis-à-vis some people but not others, depending on a broad range of factors. An individual may be reasonably likely to be identified by someone, but not have a reasonable expectation of privacy in relation to that party. Conversely, they may have a reasonable expectation of privacy in relation to someone, but not be reasonably likely to be identified by them. In other cases, reasonable expectations of both identification and privacy will apply and data protection, confidentiality and privacy rights will all be engaged.

The contextuality of identifiability and privacy rights is complex, but congruent. Understood in this way, anonymity does not draw a bright, severing line between people and their data protection rights. It is, I suggest, one of at least two key context-specific distinctions that run through data which helps to determine what kind of rights are engaged. This helps to locate relative anonymity in a broader context of information governance law, which adds weight to the characterisation as initially discussed in terms of cell and tissue donation.

VII CONCLUSION

This paper has advocated a relative, governance based understanding of anonymity.

Anonymity has been shown to be a legal standard, which requires the elimination of

reasonable means of identification. I have argued that, where ‘anonymisation’ fails, this is a failure to meet the standard of anonymity, and not a failure of the standard itself.

The relative account has been explored via a number of pieces of legislation. We have seen that Recital 26 GDPR has been interpreted by the CJEU with some relativity, and not meaning that data must be unidentifiable for every theoretical beholder to be ‘rendered anonymous.’ Crucially, I have argued that data do not have to be anonymous for a discloser to be anonymous for a recipient, as long as the recipient could not reasonably approach them for assistance in identifying people (and do not have other means of identification). The ability of a discloser to identify subjects is, I have suggested, a poor benchmark to judge identifiability for everyone else. It is often impossible for them to delete their own raw data, but this should not prevent them sharing information to the standard of anonymity, and protecting subjects from identification by a third party.

This relative perspective has been assisted by an analysis of donor anonymity and the anonymity of clinical trial subjects under the CTR. In both contexts, it appears that some individual-level information must be shared, even when it might well be identifiable for the discloser. I have argued that this is compatible with anonymity as seen from a relative perspective. However, I have also argued that anonymity should be assessed with reference to governance, and the mechanisms in place to prevent re-identifying behaviour, which points against publication of individual level information on an ‘anonymous’ basis.

Despite its limitations, the value of anonymisation lies in its capacity to assure subjects that they are not reasonably likely to be identified, and their privacy and confidentiality has therefore been preserved. It does not maintain trust in the concept to suggest anonymisation consists of techniques which open the door to unregulated uses of data, when in fact the maintenance of anonymity requires an equivalent standard of protection. While this may not

necessarily prove an easier alternative to data protection law, its purpose is to protect subjects' identities, and thus their rights to privacy and confidentiality, and not to serve controller convenience.

This account of anonymity, if accepted, naturally has implications outside the medical context. For example, the UK has recently established a system for re-use of non-medical administrative data for research under the Digital Economy Act 2017. While de-identification of data is an initial step in minimising identification risk, oversight, training and accreditation of all bodies involved in processing the information helps to eliminate likely means of identification, at least for the ultimate recipients of such information. Accredited researchers, processors and peer-reviewers must agree to be listed in a public register for transparency purposes, along with a summary of their project.⁸² Data access agreements are supplemented by a Code of Practice for data sharing, as well as a long list of counts by which researchers can lose their accreditation, including negligently facilitating the identification of individuals in the data.⁸³ The scrutiny, public profile, training and legal obligations compounded within this statutory structure provide a helpful example of governance measures which help to minimise identification risk and support relative anonymity. It is suggested that approaches such as these could be more important in a move towards a conceptualisation of anonymity grounded in information governance.

⁸² UK Statistics Authority, 'List of accredited researchers and research projects under the Research Strand of the Digital Economy Act' <<https://www.statisticsauthority.gov.uk/about-the-authority/better-useofdata-statistics-and-research/betterdataaccess-research/better-use-of-data/list-of-accredited-researchers-and-research-projects-under-the-research-strand-of-the-digital-economy-act/>> accessed 27 February 2020

⁸³ UK Statistics Authority, 'Research Code of Practice and Accreditation criteria' <https://www.statisticsauthority.gov.uk/wp-content/uploads/2018/08/COP_Research-and-Accreditation_A4.pdf> accessed 27 February 2020