Available online at www.sciencedirect.com

ScienceDirect

www.compseconline.com/publications/prodclaw.htm

Computer Law & Security Review

Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK

Miranda Mourby [a,*], *Elaine Mackey* [b], *Mark Elliot* [b], *Heather Gowans* [a], *Susan E. Wallace* [a,c], *Jessica Bell* [a], *Hannah Smith* [a], *Stergios Aidinlis* [a], *Jane Kaye* [a]

[a] *Centre for Health, Law and Emerging Technologies ('HeLEX'), Nuffield Department of Population Health, University of Oxford, UK*

[b] *School of Social Sciences, University of Manchester, UK*

[c] *Department for Health Sciences, University of Leicester, Leicester, UK*

Keywords:
Data protection
Privacy
Anonymisation
Pseudonymisation
Administrative data
Re-identification
General Data Protection Regulation

A B S T R A C T

There has naturally been a good deal of discussion of the forthcoming General Data Protection Regulation. One issue of interest to all data controllers, and of particular concern for researchers, is whether the GDPR expands the scope of personal data through the introduction of the term 'pseudonymisation' in Article 4(5). If all data which have been 'pseudonymised' in the conventional sense of the word (e.g. key-coded) are to be treated as personal data, this would have serious implications for research. Administrative data research, which is carried out on data routinely collected and held by public authorities, would be particularly affected as the sharing of de-identified data could constitute the unconsented disclosure of identifiable information.

Instead, however, we argue that the definition of pseudonymisation in Article 4(5) GDPR will not expand the category of personal data, and that there is no intention that it should do so. The definition of pseudonymisation under the GDPR is not intended to determine whether data are personal data; indeed it is clear that all data falling within this definition are personal data. Rather, it is Recital 26 and its requirement of a 'means reasonably likely to be used' which remains the relevant test as to whether data are personal. This leaves open the possibility that data which have been 'pseudonymised' in the conventional sense of key-coding can still be rendered anonymous. There may also be circumstances in which data which have undergone pseudonymisation within one organisation could be anonymous for a third party. We explain how, with reference to the data environment factors as set out in the UK Anonymisation Network's *Anonymisation Decision-Making Framework*.

* *Corresponding author.* Centre for Health, Law and Emerging Technologies ('HeLEX'), Nuffield Department of Population Health, University of Oxford, UK.

E-mail address: miranda.mourby@dph.ox.ac.uk (M. Mourby).

https://doi.org/10.1016/j.clsr.2018.01.002

The forthcoming General Data Protection Regulation ('GDPR')[1] is poised to have wide-ranging impact on those who work with data – how much impact will naturally depend on its interpretation in practice. Whether and in what circumstances de-identified data can be anonymous is an issue of great practical importance for data controllers, but one which has not escaped controversy, particularly given the ambiguity surrounding the concept of pseudonymisation.

Article 4(5) GDPR defines pseudonymisation as the processing of personal data in such a manner that they can no longer be attributed to a specific data subject without the use of additional information, with technical and organisational measures to ensure that they are not attributed to an identified or identifiable natural person. While the GDPR was in its development, some commentators predicted negative implications for research if a subset of 'pseudonymous' personal data was introduced,[2] and even after the final version has been published there appears to be a tendency to regard data as personal if they resemble data which have undergone a process of pseudonymisation.[3]

Instead, however, the GDPR defines pseudonymisation as an act of processing, and not as a category of personal data. It is therefore inadvisable to use the definition of pseudonymisation to determine whether data are personal data. We suggest that the following two-stage reasoning should be followed:

1) Are natural persons identifiable within the meaning of Recital 26, taking into account all the means reasonably likely to be used?
2) If the answer to the above question is yes, has 'pseudonymisation' been applied within the meaning of Article 4(5) GDPR?

The first section of this article explores the concepts of pseudonymisation and anonymisation under the GDPR. We will then examine the importance of anonymisation in potentially sensitive areas such as administrative data research; i.e. research undertaken using data held by public authorities in connection with their functions.[4] Finally, we will consider how anonymisation can be achieved under the GDPR, with reference to the 'data environment' factors set out in the *Anonymisation Decision-Making Framework*.[5] Anonymisation under the GDPR is, we suggest, still possible for key-coded data, and even data which have undergone pseudonymisation per Article 4(5)[6] may be anonymous when shared with a third party.

1. GDPR pseudonymisation and anonymisation

1.1. *Pseudonymisation: GDPR vs 'conventional'*

Article 4(5) GDPR defines pseudonymisation as:

> *the processing* ***of personal data*** *in such a manner that* ***the personal data*** *can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that* ***the personal data*** *are not attributed to an identified or identifiable natural person.*

As the emphasis added above illustrates, the definition evidently envisages that the data in question begin and end the process as personal data. Personal data are defined as data '*relating to*' an identified, or identifiable, data subject.[7] The data processed per Article 4(5) evidently still *relate* to an identifiable natural person; pseudonymisation merely prevents the *attribution* of the data to a natural person. In other words, GDPR pseudonymisation prevents direct identification through attribution, but not through any other means reasonably likely to be used to identify an individual, which must be excluded before he or she is no longer considered to be identifiable.[8]

The word 'pseudonymisation' in the GDPR thus refers to a process which reduces the risk of direct identification, but which does not produce anonymous data. Pseudonymisation is referred to as a means of reducing risks to data subjects,[9] and as an appropriate safeguard for any personal data used for scientific, historical or statistical research.[10] Personal data which have undergone pseudonymisation are within scope of the GDPR, and the data subject rights set out in Articles 15–20 still apply.[11]

[1] 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'.

[2] Leslie Stevens, 'The Proposed Data Protection Regulation and its Potential Impact on Social Sciences Research in the UK' [2015] EDPL 107.

[3] Matthias Berberich and Malgorzata Steiner 'Blockchain Technology and the GDPR – How to Reconcile Privacy and Distributed Ledgers?' [2016] EDPL 424.

[4] This definition of 'administrative data' is taken from s.64 Digital Economy Act 2017, which provides new powers of disclosure for public interest research.

[5] Mark Elliot, Elaine Mackey, Kieron O'Hara and Caroline Tudor, *The Anonymisation Decision-Making Framework* (UKAN, 2016).

[6] The GDPR does not use the word 'pseudonymous' or 'pseudonymised', although the word 'pseudonymised' has been used by the Article 29 Working Party in their Guidance WP260 on Transparency under the GDPR. For the most part we will refer in this paper to 'data which have undergone a process of pseudonymisation', or similar. If, for ease of expression, the term 'GDPR pseudonymised data' is used in this paper, it is only as a shorthand for 'data which have undergone a process of pseudonymisation'.

[7] GDPR, Article 4(1).

[8] GDPR, Recital 26, as discussed in more detail in section 2.3.

[9] GDPR, Recital 28.

[10] Article 89 & Recital 156.

[11] It is possible, however, that use of pseudonymised data may fall within Article 11 GDPR – processing in which it is not necessary to identify the data subject – in which case these data subject rights may not apply, see Article 29 Working Party *Guidelines of transparency under Regulation 2016/679* WP260, para 57.

The GDPR definition of pseudonymisation differs significantly from the conventional way in which the term has been used. For example, the Anonymisation Decision-Making Framework defines pseudonymisation as:

> *A technique where direct identifiers are replaced with a fictitious name or code that is unique to an individual but does not itself directly identify them.*[12]

Similarly, the Information Commissioner's Office defines pseudonymisation as:

> *The process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity.*[13]

These orthodox articulations of pseudonymisation fall short of what is required within pseudonymisation under the GDPR. The GDPR does not merely describe a technique or a process – in fact, it does not specify at all what techniques should be used, other than stating that a 'process' must be applied. GDPR pseudonymisation requires not just a process but an ultimate 'success state', in which the data cannot be attributed to an individual without the use of additional information. Even this additional information is addressed within the definition, as it must be subject to '*technical and organisational measures*' to prevent reattribution. Thus, the data must not only be modified so that they are not directly identifiable, they must also be protected against re-identification.

The significance of this discrepancy is that conventional characterisations of pseudonymisation are neutral as to whether the resulting data are personal or anonymous; whereas under the GDPR, the pseudonymised data are personal, but protected against identification. This raises the question of whether 'pseudonymisation' can still be discussed as part of an anonymisation process, or whether all 'pseudonymised data' must be considered personal once the GDPR is in force.

1.2. *Pseudonymisation as anonymisation?*

In 2014, the EU Article 29 Working Party produced guidance to the effect that:

> *pseudonymisation is not a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure.*[14]

This is, in essence, the meaning of 'pseudonymisation' within the GDPR. In combination with the stipulation elsewhere in the 2014 guidance that de-identification must be 'irreversible' to be considered 'anonymisation', it appears to form a strong case against the use of pseudonymisation to create anonymous data, especially where original identifying information is retained.

It is trite to say that data which have undergone pseudonymisation in the GDPR sense are personal data; this is merely a function of the definition given in Article 4(5). However, as noted above, the word 'pseudonymisation' is not always afforded the same meaning. The UK Information Commissioner's Office ('ICO') suggests that pseudonymisation *can* produce anonymised data on an individual-level basis.[15] While it acknowledges that this data may pose a greater privacy risk than aggregate anonymous data, they ultimately conclude:

> *This does not mean though, that effective anonymisation through pseudonymisation becomes impossible.*[16]

The ICO has maintained this position in relation to the GDPR, advising on its website:

> *Personal data that has been pseudonymised – e.g. key-coded – can fall within the scope of the GDPR depending on how difficult it is to attribute the pseudonym to a particular individual.*[17]

This is to say, data which have been pseudonymised *can* fall within the scope of the GDPR, i.e. they *can* be personal data, but this is not necessarily the case. Conversely, if they *can* fall outside the scope of the GDPR, then it must follow that they can be anonymous. 'Pseudonymised' in this context evidently means 'key-coded,' or the equivalent. This guidance may contrast with the definition of pseudonymisation within the GDPR, but makes sense if the word 'pseudonymised' is understood merely as a de-identification technique for individual level data – or the 'conventional sense' as it is referred to in this paper.

The ICO's guidance on the status of pseudonymised data under the GDPR has been adopted by the Government in the Explanatory Notes to the current Data Protection Bill,[18] and will be a key point of reference for UK data controllers in determining whether they are handling personal data. It appears from the ICO's guidance that pseudonymisation in the 'conventional' sense can still be spoken of as creating anonymised data. But what about pseudonymisation in the GDPR sense? Can data which have undergone this type of processing ever be deemed anonymous?

---

[12] Note 5, page 15.

[13] Information Commissioner's Office, *Anonymisation: managing data protection risk code of practice* (Wilmslow, November 2012) < https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 7 December 2017.

[14] Article 29 Working Party *Opinion 05/2014 on Anonymisation Techniques* WP216 (Brussels, 10 April 2014).

[15] Note 13, page 7.

[16] Ibid, page 21.

[17] ICO, 'Overview of the General Data Protection Regulation (GDPR)', https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/ accessed 6 November 2017.

[18] Explanatory Notes to the Data Protection Bill 2017-19 <https://publications.parliament.uk/pa/bills/lbill/2017-2019/0066/18066en.pdf> accessed 20 September 2017.

Public Authority A provides administrative personal data to a Research Centre, B, to be used for research purposes. The Research Centre wishes to share this data with Researcher C, but is not sure whether they would be disclosing personal data.
Research Centre B processes the personal data, and removes the information which is deemed to be directly identifying. These identifiers are held separately within Research Centre B, with technical and organisational controls to prevent their reattribution to the research data.
Researcher C accesses the research data in a secure lab (Research Centre B). She has completed the Centre's accreditation training so she knows she cannot bring a phone or tablet into the room where she is working on the data, and the computer she works on is not networked. In addition, she signs an agreement with Research Centre B not to attempt to identify any natural person within the data (she is interested solely in the patterns within the data, which might help their project). All her analytical outputs from her work are checked before she is allowed to take them out of the centre.
Researcher C has no relationship with Research Centre B, or with Public Authority A, which would enable her to access any potentially identifying information. She has no means by which she is reasonably likely (or indeed likely) to obtain the information which would identify the data. The information exists, and so identification is not theoretically impossible and the processing is therefore not technically irreversible. However, it is extremely unlikely that the researcher could or would have access to any information which would enable her to identify natural persons within the data.
Is Researcher C accessing personal data?

The example set out in the Box above illustrates the ambiguity of the relationship between pseudonymisation and anonymisation. Following the 2014 Article 29 Working Party guidance, it would be possible to conclude that the researcher is accessing personal data, as the de-identification processing would not be irreversible, just very unlikely to be reversed by the researcher. To use the definition of pseudonymisation under Article 4(5) as a benchmark would also yield the conclusion that these data are personal; that is simply to say they have undergone a process of pseudonymisation and therefore they are personal.

However, it is the argument of this paper that a more nuanced approach should be adopted. Following the logic of the Court of Justice of the European Union in *Breyer v Germany*, the focus should be on the relationship between the parties, and whether these relationships enable the researcher to identify the data. As explained in more detail below, *Breyer* is authority for the proposition that the scope of personal data should be determined by assessing whether there is a means reasonably likely to be used to identify individuals, and not merely a theoretical possibility of identification.

As it is Recital 26 of the GDPR, and not Article 4(5), which determines whether the data are personal data, this leaves open the possibility that data which have undergone GDPR pseudonymisation could be anonymous for a third party such as Researcher C.

In short, 'conventional' pseudonymisation should still be available as an anonymisation technique under the GDPR, however for the reasons outlined in the previous section it may preferably be described by an alternative term, such as 'de-identification', in order to delineate it from the GDPR definition.

As to whether anonymisation is possible for GDPR pseudonymised data,[19] it is necessary to consider the case of *Breyer* in more detail.

1.3. *GDPR anonymisation:* Breyer v Germany

To understand when data can be considered 'rendered anonymous' under the GDPR, it is necessary to consider the detail of Recital 26. In its final form, Recital 26 GDPR reads as follows:

> *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. 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. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.*

While its length and detail do not necessarily render it immediately accessible, the correct interpretation of Recital 26 is essential to understand the rest of the GDPR. It sets out the test to determine whether a natural person is identifiable, and therefore whether any data are personal and thus within scope of the Regulation. At the beginning of this paper, we suggested that the question of whether data are personal should be separated from that of the application of pseudonymisation; this is to prevent any confusion which might result from the reference to pseudonymisation within Recital 26.

Prior to the finalisation of the GDPR, the ICO warned of the possible confusion stemming from the text of this Recital. Writing on the Council's text of what was then Recital 23, the Commissioner's Office commented as follows:

> *In our view there should be a single definition of 'personal data'. Therefore it is welcome that 'pseudonymous data' is no longer treated as a separate category of personal data. However, pseudonymisation should only be relevant as a privacy enhancing technique – for example in relation to data minimisation or security. It would be better not to try to define pseudonymisation in the context of the definition of personal data.*
>
> *As it stands, the relevant Recital (23) is confusing. It says that pseudonymous data should be considered as information on an*

[19] See note 6 above.

> *identifiable natural person – this implies all pseudonymous data whoever it is held by. However, the relevant Recital's new reference to the likelihood of identification presumably means that some pseudonymous data would be personal data whilst other pseudonymous data would not be, depending on the likelihood of the relevant identifying information being added to the pseudonymous information.*[20]

Unfortunately, despite the ICO's warning, discussion of pseudonymisation *has* been included in Recital 26. This gives the impression that pseudonymisation is determinative of identifiability, rather than a process to be applied to personal data as defined in Article 4(5). This in turn creates confusion between the test to determine whether data are personal, and the definition of whether personal data have been successfully pseudonymised.

For example, the ICO's guidance that key-coded data and IP addresses 'can' fall within the scope of the GDPR contrasts with Matthias Berberich and Malgorzata Steiner's analysis of Blockchain ('BC') data under the GDPR. Citing the then pending case of *Breyer v Germany*,[21] they reason as follows:

> *Whether the use of BC must comply with the GDPR, will first and foremost depend on whether personal data is stored on BC under respective business model. Most of currently discussed use cases involve transactions of all sorts, be it financial assets, property registers or smart contracts, which all usually involve specific information in relation to a specific person as a rightholder, owner or contract party. Albeit this information is normally encrypted and can only be accessed with the correct keys, encryption of the data as such- i.e. giving access only to authorised parties – will normally not take such information out of the scope of the GDPR. Even if personal information only entails reference ID numbers, such identifiers are typically unique to a specific person. While in all such cases additional information may be necessary to attribute information to the data subject, such information would be merely pseudonymised and count as personal information.*[22]

Under this analysis, encrypted data would count as data which have undergone pseudonymisation, because the process of encryption appears to correlate with the process of pseudonymisation under Article 4(5). However, if the two-stage process advocated in this paper were adopted, the relevant questions would be:

1) Can the data held on Blockchain identify natural persons by any means reasonably likely to be used?
2) If so, have they undergone pseudonymisation within the meaning of Article 4(5)?

The answers to these questions will inevitably vary from one situation to the next, but if the data are found to be personal data this should be because they could identify individuals through means 'reasonably likely to be used,' and not because the encryption process resembles pseudonymisation.

The importance of considering such means as are 'reasonably likely' to be used was highlighted by the Court of Justice of the European Union in their ultimate judgment in the *Breyer* case. The case brought by Patrick Breyer against the German Government was referred to the Court of Justice of the European Union ('CJEU') for a determination as to whether the data in question were personal data. The German government held the IP addresses of individuals who had visited public authority websites; while these addresses related to natural persons, they could not be attributed to identifiable individuals without the use of additional information. This additional information was held by Internet Service Providers.

The data in question were not pseudonymised or de-identified; rather, they were partial, and could only be identified by the additional information. In this respect they were analogous to pseudonymised data, although the question posed was whether individuals could be identified through construction, not reconstruction, of a dataset. The debate as to whether the data were personal, therefore, mirrors the debate over whether pseudonymised data are always personal data.

The opinion of the Advocate General essentially corresponds to the argument that pseudonymised data are always personal. It was submitted by AG Sanches-Borodona that data would be personal as long as a known third party held identifying information, which *could* be used to identify the data (regardless of likelihood of attribution). The opinion was criticised by some on the grounds that it represented an absolute approach which would extend the scope of the GDPR too widely, burdening data processing entities in a way which would be incommensurate with the actual risks to the privacy of data subjects.[23]

The opinion of the Advocate General was not, in the event, followed by the CJEU, who favoured what was termed a more 'relative' approach. The CJEU favoured a logic under which the *means* by which this additional data could be used to identify the data were taken into account. In their view, it was necessary to determine '*whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means likely reasonably to be used to identify the data subject.*'[24] They concluded that the data were personal, but only because of the existence of legal channels enabling the competent authority to obtain identifying information from the internet service provider in the event of a cyber-attack.[25] In the absence of these channels, the data would not have been considered personal simply because a known third party could identify them:

> *Thus it appears that the online media services provider has the means which may likely reasonably be used in order to identify*

[20] 'ICO analysis of the Council of the European Union text of the General Data Protection Regulation', <https://ico.org.uk/media/1432420/ico-analysis-of-the-council-of-the-european-union-text.pdf> accessed 6 November 2017.

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

[22] Matthias Berberich and Malgorzata Steiner 'Blockchain Technology and the GDPR – How to Reconcile Privacy and Distributed Ledgers?' [2016] EDPL 424.

[23] Gerard Spindler and Philipp Schmechel, 'Personal Data and Encryption in the European Data Protection Regulation' [2016] JIPTEC 7 (2).

[24] Note 21 at 45.

[25] Ibid at 47.

the data subject [. . .] a dynamic IP address registered by an online media service provider [. . .] constitutes personal data [. . .] in relation to that provider, where the latter has the legal means which enable it to identify the data subject with additional data which the internet service provider has about that person.[26]

The judgment of the CJEU in *Breyer* is crucial authority on the interpretation of Recital 26 of the Directive, and, by extension, Recital 26 GDPR. While the IP addresses in question were not pseudonymised, but partial, data, the principles on which they were judged to be personal closely mirror those which would apply to pseudonymised data. Unless a drastically different approach is applied to individual level, de-identified data under the GDPR, it should be possible for these data to be personal or anonymous, depending on the circumstances.

While *Breyer* was, of course, a judgment on the existing Directive and not the GDPR, the text of Recital 26 in the GDPR and in the Directive differ very little in substance. The relevant sentence in the Directive reads:

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 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.

The inversion of 'reasonably likely' from 'likely reasonably' in the GDPR does not alter the meaning of the sentence, although it is perhaps more natural English syntax. The only additions to the Recital 26 in the GDPR appear to be the explicit mention of singling out as a method to be considered as a potential means of identification (of which, more below in section 2.3). To introduce a category of data which are always identifiable, irrespective of likelihood of attribution, would represent a significant departure from the Directive, for which one would hope for clearer evidence. The fact that references to 'pseudonymous data', as proposed in 2013,[27] are not present in the final version of the GDPR, suggests the decision was taken not to introduce such a category.

Therefore, if the precedent set by *Breyer* is to be applied to data which have undergone pseudonymisation under the GDPR, it should be possible for these data to be rendered anonymous in some circumstances. To return to the example given in section 2.2 of GDPR pseudonymised data held by a research centre, and then shared with an external researcher ('Researcher C'), these shared data **would not be personal data** for Researcher C. The key point in *Breyer* is whether the relationship between the parties is such that the researcher has any means reasonably likely to be used to identify data subjects. While these data are undeniably personal for the research centre, they will not be personal for the researcher if she has no means reasonably likely to be used to access the identifiers.

In short, even data which have undergone a process of GDPR pseudonymisation may not be personal data when shared with third parties. It is worth noting that this interpretation is consistent not only with the judgment in *Breyer*, but with the view of anonymisation the UK Government has enacted in the Digital Economy Act 2017. Once s.64 of this Act comes into force, the researcher in the example could access this information as long as it is processed prior to disclosure so that:

a) no person's identity is specified in the information, and
b) it is not reasonably likely that any person's identity will be deduced from the information, by itself or together with other information.[28]

This suggests the UK Government is satisfied that 'good enough' anonymisation can be achieved for the disclosure of administrative data, even when the processing is theoretically not irreversible and original identifiers are retained by the data controller. Analysis provided by the Wellcome Trust also supports this position, and coincides with the ICO response to the draft Recital 23 cited at the beginning of this subsection:

Recital 26 can be read that all pseudonymised data should be considered personal data. . . However, the scope of identifiability is qualified by the reference to "means reasonably likely to be used" as under the 1995 Directive. This suggests that there may be cases where pseudonymised data together with a combination of appropriate organisational, legal and technological measures can be considered anonymous data. A proportionate and context-dependent approach would take into account the range of measures used, including pseudonymisation, to determine whether the data is considered to be identifiable. In order to achieve this it is important to consider the text of Recital 26 in full to understand how the scope of the regulation relates to approaches commonly used in research.[29]

Applying this argument in a research context, therefore, even if data are personal and GDPR pseudonymised within a research centre, it is possible for them to be anonymous for a third party researcher.

1.4. *Singling out*

The other respect in which it has been suggested that the GDPR expands the scope of personal data is through reference to 'singling out' in Recital 26. It has been suggested by the authors of the authoritative *Companion to Data Protection Law and Practice* that it is clear that as long as a person can be singled out he or she is regarded as identifiable.[30] As Leslie Stevens argues, if this is the case it would have serious consequences for administrative data research.[31]

Individuals within pseudonymised data can be singled out, in the sense of being distinguished from one another, and as such the question of singling out is relevant for our present

[26] Ibid at 48–49.
[27] Note 20.
[28] Digital Economy Act 2017, s.64(3).
[29] Beth Thompson, 'Analysis: Research and the General Data Protection Regulation,' Wellcome Trust July 2016, page 2 https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf.
[30] Rosemary Jay, *Guide to the General Data Protection Regulation: A Companion to Data Protection Law and Practice* (Sweet & Maxwell 2017) page 339.
[31] Note 2.

purposes. The phrase 'singling out' as used in the GDPR appears to originate from the Article 29 Working Party's 2007 guidance on personal data,[32] although this was built upon previous discussion of 'piecing together' an individual's identity through various pieces of information.[33] Discussion of singling out has led some to suggest that tracking cookies constitute personal data,[34] although this may not be consistent with the CJEU's treatment of IP addresses in *Breyer*.

In her analysis of the proposed drafts of the GDPR, Stevens identifies a possible interpretation under which data in which individuals can be singled out (if not directly identified) are personal data. This interpretation stems from an earlier draft of Recital 26 (or Recital 23 as it then was). The EU Parliamentary draft of Recital 23 suggested that any individual capable of being 'singled out' is identifiable, and his or her data is therefore personal:

> *To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used either by the controller or by any other person* ***to identify or single out*** *the individual directly or indirectly (emphasis added).*

As Stevens points out, it is clear from this draft that an individual is identifiable as long as he or she is capable of being singled out, as the terms 'to identify' and 'to single out' a person are treated as equivalent. In the final text of the GDPR, however, it appears that singling out is only referred to as a method for identification:

> *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 may be subtle, but the movement of the phrase 'singling out' from the end result ('identification or singling out') to one of the methods to be used to achieve the end result ('singling out to identify') dramatically changes the meaning of the Recital. In the final version of Recital 26, therefore, it is clear (if perhaps tautological) that for a person to be identifiable, it must be reasonably likely they will be identified. Singling out need only be considered if it is a means 'reasonably likely' to be used to achieve this end.

Recital 26 explains that the reasonable likelihood of a particular method being used to identify an individual should be determined in light of all the 'objective factors', including cost of and time required for identification, and the available technology. No one method of identifying an individual is considered 'reasonably likely' to identify individuals in all cases, each set of data must be considered in its own unique set of circumstances. Under this case-specific approach, it appears entirely possibly for de-identified, individual-level data to be rendered anonymous as long as the reasonably likely means of identification have been eliminated. Therefore, the fact that individuals can be singled out (i.e. individually differentiated) within 'pseudonymised data' is not sufficient to render these data personal.

One caveat is worth noting, however. It appears from guidance issued by the ICO[35] in relation to the Data Protection Act 1998 that if a public authority tracks an otherwise unnamed and unknown individual (for example, through CCTV cameras) this is sufficient identification for data protection principles to apply. Similarly, it has been argued that the use of online data to profile or target individuals should constitute use of personal data, even if the individuals in question are unnamed.[36] It may therefore be necessary, if administrative data are to be considered anonymous when used by a researcher, to ensure through safeguards that the researcher does not profile any individual within the data.

1.5. *Summary*

The above subsections have explored how 'conventionally pseudonymised' data can still be 'rendered anonymous' for the purposes of the GDPR, how even data which have undergone GDPR pseudonymisation could be legally anonymous if shared with a third party, and how 'singling out' is a *method* of identification, and not identification itself. These three points are vital to rebutting the suggestion that the GDPR increases the scope of personal data to the detriment of research.

The next section addresses *why* an increase in the scope of personal data would be to the detriment of research, with a particular focus on research conducted using administrative data.

2. Anonymisation of administrative data

Administrative data research is one of many areas where clarity as to whether pseudonymised data are personal or anonymous is essential. 'Administrative data' is a term used to describe the information held by public authorities in connection with their functions. Since the report of the Administrative Data Taskforce in 2012,[37] it has been argued that the UK should do more to exploit the secondary usage of this type of data, which can provide vital evidence of the impact of government policy and help shape future policy-making for the better. This type of research, usually carried out using the pseudonymised microdata of those who use UK public services, must avoid infringing the privacy or confidentiality of

[32] Article 29 Working Party, *Opinion 4/2007 on the concept of personal data* WP136 (Brussels, 20 June 2007).

[33] Council of Europe's T-PD Committee, *Report on the application of Data protection principles of the worldwide telecommunication network*, ((2004) 04 final), point 2.3.1.

[34] Frederik J. Zuiderveen Borgesius, 'Singling out people without knowing their names – Behavioural targeting, pseudonymous data, and the new Data Protection Regulation' [2016] C.L.S. Rev 256.

[35] Information Commissioner's Office, *Determining what is personal data* (Wilmslow, 2012) <https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf> accessed 3 January 2018, page 8.

[36] Note 34.

[37] *The UK Administrative Data Research Network: Improving Access for Research and Policy*, Report from the Administrative Data Taskforce, December 2012, <www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-administrativedatataskforcereportdecember201_tcm97-43887.pdf> accessed 11 October 2017.

data subjects. Anonymisation is thus a key means of ensuring that such research is conducted legally, ethically, and in such a way that the public would support.[38]

If anonymisation of individual-level administrative data is not possible under the GDPR, the regulatory burdens on research centres will increase, and the publication of findings may be impeded. Crucially, anonymisation could also lose its effectiveness in ensuring there is no breach of common law confidentiality. The distinction between personal and anonymous data has long been used as a benchmark for whether a duty of confidentiality arises,[39] or whether rights under Article 8 European Convention on Human Rights ('ECHR')[40] are engaged.[41] If pseudonymised data can no longer be considered anonymous, the pool of valuable data which remain legally 'safe' to work with could drastically shrink, with commensurate damage to this developing field of public interest research.

2.1. *Confidentiality*

The first and perhaps most important reason why anonymisation is desirable is the duty of confidentiality owed by public authority to the people whose data they hold. This is a duty owed under common law, which is classically formulated as follows: is the information itself of a confidential nature (i.e. is it publicly available?), is it communicated in circumstances importing an obligation of confidence, and has that confidence been breached?[42]

Administrative data is a broad category, encompassing many types of information which may be confidential, and stemming from a range of different relationships which might attract confidentiality of communication (for example, a doctor–patient relationship).[43] Information has been held to be confidential if it relates to family[44] and intimate relationships[45]; some administrative data will inevitably contain such information.

As a general rule, the Government has said it is likely that names and addresses of individuals supplied to public bodies in pursuance of their functions would in some cases amount to confidential information.[46] Individuals can reasonably expect their public sector personal data will be held confidential unless they are given some prior warning to the contrary.[47] Therefore, while there may be exceptions, it is safest to assume that administrative data will attract a duty of confidence. Data are only deemed to be confidential if they contain information about identifiable individuals[48]; identifiability being determined in accordance with data protection law.

Consequently, if data have been anonymised a public authority can justifiably claim that there has been no breach of confidentiality as the data have not been passed onto researchers in an identifiable form. The Court of Appeal's judgment in *R (Source Informatics) v Department of Health*[49] still stands: disclosure of anonymised data to a third party does not constitute a breach of confidence under UK law. Anonymisation thus prevents breach of confidentiality when public sector data are provided to a third party such as a researcher. Conversely, if pseudonymised microdata were inevitably deemed to relate to identifiable individuals, the accusation could be made that sharing these data with researchers constitutes a breach of confidence, even if the researchers had no practical power or motivation to identify the data subjects.

A well-known example of administrative data research gone wrong is *care.data*, where the public backlash was triggered in part by the fear that pseudonymisation alone would not be enough to protect patient identities, particularly as commercial purchasers of NHS data might have the resources to re-identify individual patients.[50] In light of these concerns, GDPR pseudonymisation affords insufficient protection for administrative research data. For example, Latanya Sweeney successfully re-identified the pseudonymised health records of the then Governor of Massachusetts by cross-referencing with publicly available aggregate census data,[51] rather than using his original health records.

By contrast, the broader considerations within GDPR anonymisation would include questions such as: who has access to the data in question, what other information would they have access to, for what purpose are they using the data, and what resources or motive might they have to try to identify data subjects? All of these factors would inform whether individuals can be identified by means reasonably likely to be used, and could equally be of concern to the public in terms of how their data are protected. In this context, therefore, GDPR anonymisation is the more appropriate option.

2.2. *Privacy*

The right to privacy under Article 8 of the ECHR imposes upon the State an obligation to respect the private lives of the individuals with which it deals.[52] This means that any interference with citizens' rights must be proportionate within the meaning of the ECHR. Any public authorities disclosing administrative data for research will therefore need to do so in a way which is compatible with this obligation.

It is possible that, where data have been anonymised, there will be no interference with ECHR privacy rights. It has been

[38] Ibid at 38.
[39] *R (Source Informatics) v Department of Health* [2001] Q.B 424.
[40] Convention for the Protection of Human Rights and Fundamental Freedoms, Rome, 1950 Council of Europe European Treaty Series 5.
[41] *South Lanarkshire Council v The Scottish Information Commissioner* [2013] UKSC 55 at 26.
[42] *Coco v AN Clark* [1969] RPC 41.
[43] Sir Roger Toulson and Charles Phipps, *Confidentiality* (3rd ed., Sweet & Maxwell 2012) at 11001.
[44] *Argyll v Argyll* [197] Ch 302.
[45] *Lennon v News Group* [1978].
[46] Department for Constitutional Affairs, *Public Sector Data Sharing: Guidance on the Law* (DCA 2003) page 20.
[47] *W, X, Y & Z v Secretary of State for Health* [2015] EWCA Civ 1034.
[48] See note 49 below.
[49] [2001] Q.B 424.
[50] Lizzie Presser, Maia Hruskova, Helen Rowbottom and Jesse Kancir, 'Care.data and access to UK health records: patient privacy and public trust' (2015) Technology Science <https://techscience.org/a/2015081103/> accessed 11 October 2017.
[51] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', (2010) 57 UCLA Law Review 1701, 1719.
[52] *Von Hanover v Germany (2005) 40 E.H.H.R.* 1, at 57.

suggested by the Supreme Court in *South Lanarkshire Council v The Scottish Information Commissioner*[53] that if disclosure '*would not enable* [the data recipient] *or anyone else to discover the identity of the data subjects,*' it would be '*quite difficult to see why there is any interference with their right to respect for their private lives.*'[54] This comment may be predicated on the 'firm distinction' between identifiable and non-identifiable data which Aldhouse and others disavow[55] (can one be certain that an unspecified 'anyone else' could not identify an individual's data?). Yet it is, nonetheless, an indication of the usefulness of using anonymised data to protect privacy.

2.3. Criminal offences

Among the benefits listed by the ICO in complying with their Anonymisation Code of Practice is '*minimising the risk of breaking the law and consequent enforcement action by the Information Commissioner's Office (ICO) or other regulators.*'[56] Under s.66 of the Digital Economy Act 2017 it is a criminal offence to disclose any personal information from administrative data received for research purposes. Such activity would already be an offence under s.55 Data Protection Act 1998 if undertaken without the data controller's consent. Under the Data Protection Bill, it will be a criminal offence to re-identify de-identified data without the consent of the data controller.[57] Again, pseudonymisation itself is not enough to minimise the risk of committing offences under these Acts, but governance of who has access to the data, for what purposes, and in what conditions better mitigate against any potential incursion of criminal liability.

2.4. Data protection

Under the current UK data protection law, set out in the Data Protection Act 1998, pseudonymised data are like any other type of data, and are subject to the data protection principles unless they do not identify a living individual, either on their own or in conjunction with other information which is (or is likely to be) in the possession of the data controller.[58] The ICO views pseudonymisation as one technique capable of producing anonymised data,[59] but does acknowledge that pseudonymised data may be particularly vulnerable to identifying individuals through the 'matching' of various pieces of 'anonymised' information.[60] Under the GDPR, data which have undergone pseudonymisation may be exempt from certain data subject rights, such as subject access, correction, erasure and data portability requests, as long as the controllers can demonstrate that they themselves are not in a position to identify the data subject.[61] However, while the data remain personal, the relevant data controller will still have a duty to keep records of processing, to carry out privacy impact assessments, to appoint a Data Protection Officer and to demonstrate compliance with the principle of privacy by design.[62] It is arguable whether these data protection obligations are in practice any less burdensome than the requirements of maintaining anonymity under the GDPR.

2.5. Summary

There would be a number of negative implications for administrative data research if the scope of personal data was increased by the GDPR: potential breaches of confidentiality, greater infringement of privacy and increased regulatory obligations. Additionally, if anonymisation were to be supplanted by pseudonymisation as defined in Article 4(5) GDPR, this would restrict the consideration of identification risk to the narrow set of 'technical and organisational measures' required by the definition. This argument is explored further in the next subsection.

2.6. Anonymisation vs pseudonymisation

To return to the definition of pseudonymisation – it is clear that not only is it a very specific definition, but it also requires a very specific form of data protection:

> *The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.*

While this requires '*technical and organisational measures*', it is the '*additional information*' which must be '*subject*' to these measures. It is thus the additional identifiable information, held separately from the pseudonymised data, which must be protected. Therefore, the only risk of identification mitigated against within GDPR pseudonymisation is the risk of identification through the original data held by the controller (or by a third party). Fuller clarification of this definition of pseudonymised data can be found in Recital 29 GDPR:

> *In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organisational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately (emphasis added).*

Reading the two provisions together, it is clear that the only protection required for pseudonymised data is against what

[53] [2013] UKSC 55.
[54] Ibid at 26.
[55] Francis Aldhouse, 'Anonymisation of personal data: a missed opportunity for the European Commission' [2014] C.L.S.Rev. 403.
[56] Note 13.
[57] Data Protection HL Bill (2017-19) 66, cl 162.
[58] Data Protection Act 1998, s.1(1).
[59] Note 13 at 7.
[60] Ibid at 21.
[61] GDPR, Article 11.
[62] GDPR, Articles 30, 35, 36, 37 and 25.

could be termed the 'internal' risk of identification from additional information retained by the data controller, or by a known third party. It is understandable that only this limited protection is expected for pseudonymised data; unless they are also 'rendered anonymous', pseudonymised data are still treated as identifiable and within the scope of the GDPR[63] and are subject to corresponding (if potentially modified) data subject rights.[64] It is logical, and consistent with existing legislation, to require broader protection before data are considered out of the scope of data protection principles.

The GDPR acknowledges that pseudonymisation is not a complete solution to issues of data protection, clarifying that its explicit introduction in the Regulation is not intended to preclude any other means of data protection.[65] Pseudonymisation alone may be sufficient where it is acceptable for the data to remain identifiable within the meaning of the GDPR; there may even be circumstances in which it is positively desirable to retain the option of re-identifying the data at a later date.[66] However, there will also be circumstances in which data are used on the understanding that the risk of identification will be minimised. The overall risk of identification is not necessarily limited to the risk posed by the original identifying data, but to any means 'reasonably likely' to be used to identify the data, in which case GDPR anonymisation becomes pertinent.

Under GDPR anonymisation, it must be determined whether natural persons are identifiable by any means 'reasonably likely' to be used. These means may include identifying the data through known additional information which is held separately, but may also comprise more indirect methods of identification, such as singling out individuals and determining their identity using other, possibly multiple, sources of information. For example, following the AOL search history disclosure in 2006, searcher no. 4417749 was identified not through the original identifying information held by AOL, but through the content of her searches which revealed her surname, age and geographical area.[67] Whether such detail in pseudonymised data could identify an individual, especially in connection with publicly available information, is a prime example of the additional considerations required as part of the process of GDPR anonymisation.

Anonymisation is thus appropriate in circumstances sufficiently sensitive to warrant broader consideration of identification risk, such as the large-scale use of public sector data for research. Pseudonymisation within the meaning of GDPR Article 4(5) would be inadequate to address all of the risks of identification encompassed within the 'means reasonably likely to be used' test. The breadth and variety of these considerations, and how they can contribute to an assessment of identification risk, are considered in the next section.

3. The data environment

We have, in the previous section, referred to the 'broader considerations' which must be taken into account as part of the process of anonymisation under the GDPR, as opposed to GDPR pseudonymisation. A more detailed and systematic review of these considerations can be found within the Anonymisation Decision-Making Framework ('the ADF'), published by the UK Anonymisation Network.[68]

The ADF is underpinned by the 'data environment' perspective, which shifts the focus from the data alone to the relationship between data and their environment.[69] The data environment is the context for any piece of data, and is considered to be made up of four components:

- Other data: the key question is what (other) data exists in a given data environment? This is what the data controller needs to know in order to assess (some of the conditions for) re-identification risk. Other data consists of co-present databases, data derived from personal knowledge and publically available sources such as public registers, social media profiles etc.
- Agency: there is no re-identification risk without human agency – this may seem like an obvious point but it is one that is often overlooked and the least understood.[70] The key issue here is in understanding who the key protagonists are, and how they might act and interact to increase or negate the risk of re-identification.
- Governance processes: these processes are formally determined in policies and procedures which control how data are managed and, accessed, by whom, how and for what purposes.
- Infrastructure: infrastructure is not dissimilar to governance, it shapes the interaction between data and environment and includes such as operational and management structures as well as hardware and software systems for maintaining security.

Although it is not within the scope of this paper to provide an in-depth analysis of the data environment perspective underpinning the ADF, we will illustrate the way it can help determine the identifiability of data. For this purpose, we apply the four data environment components to the case of the Administrative Data Research Network ('the ADRN').

[63] GDPR, Recital 26.
[64] Note 11.
[65] GDPR, Recital 28.
[66] Luca Bolognini and Camilla Bistolfi, 'Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from Directive 95/46/EC to the new EU General Data Protection Regulation' [2017] C.L.S.Rev. 171.
[67] Michael Barbaro and Tom Zeller Jr., 'A Face Is Exposed for AOL Searcher No.4417749', *The New York Times* (New York, 9th August 2006), as cited by Francis Aldhouse, 'Anonymisation of personal data – A missed opportunity for the European Commission' (note 55).
[68] Note 5.
[69] For more detail see Elaine Mackey and Mark Elliot, 'Understanding the data environment' (2013) 20(1) XRDS 36; Mark Elliot and Elaine Mackey, 'The Social Data Environment' in Kieron O'Hara, M-H. Carolyn Nguyen and Peter Haynes (eds), Digital Enlightenment Yearbook: Social Networks and Social Machines, Surveillance and Empowerment (IOS Press 2014).
[70] Elaine Mackey, 'A Framework for Understanding Statistical Disclosure Processes: A Case Study using the UK's Neighbourhood Statistics' (PhD Thesis, University of Manchester 2009).

The ADRN was designed to be a new legal, safe and efficient pathway for researchers to access linked administrative data for social and economic research. It has secure Research Centres in each of the four countries of the UK. The ADRN provides a useful case study for illustrating how accredited researchers are able to access detailed and sensitive data that are effectively anonymised, or rather what the ADF refers to as *functionally anonymised*.[71] The key point is that the ADRN employs controls on both data and environment to achieve functional anonymisation. Let us consider this in more detail.

Data received by the Research Centre are de-identified (i.e. direct identifiers are removed). Specially trained staff at the Centre prepare the data for the researcher who will access it in a secure room using a non-networked computer. The data that the researcher accesses are considered functionally anonymised. This is because a combination of data and environment controls are enacted to ensure that the re-identification risk is remote. The re-identification risk can be analysed in the context of the four data environment components outlined above: other data, agency, governance processes and infrastructure.

In terms of other data: this may relate to personal knowledge and or external information sources. The ADRN's governance processes and infrastructure around data security, access and use is such that the researcher is unable to bring into the secure environment (other) data. A researcher cannot take any materials in or out of the room in which they access the research data (including mobile phones, memory sticks or pen and paper) and they cannot copy or download data.[72] In addition, all research outputs are checked prior to being released from the Research Centre to ensure that data subjects cannot be re-identified and information about them cannot be disclosed.

In terms of agency: the key agents in this situation are the researchers. The way in which they access the data (as described above) and how they behave in the Research Centre is shaped by the ADRN's governance processes and security infrastructure. A researcher's project proposal is assessed by an independent approvals panel comprising of academics, data providers and members of the public. The panel considers whether the project is of public benefit, scientific merit and feasible. It also reviews the project's Privacy Impact Assessment which is undertaken by the ADRN If the panel approves the project those members of the project's research team who plan to access the data are required to undertake and pass researcher training prior to working in the Research Centre. The training covers: (i) legal issues, (ii) the researcher's responsibilities under the law, (iii) appropriate use of the secure setting and sanctions for misuse and (iv) processes for checking that output are not disclosive.[73] The ADRN controls the who and how of access, ensuring that those accessing the data do so safely, lawfully and responsibly.[74]

Governance processes and infrastructure inform who accesses data, and on what terms. These processes shape users' relationship with data through formal means such as policies and agreements, as well as more informal measures such as de facto norms. The governance provided by the ADRN promotes a culture of responsible use of data, and respect for legal sources of governance such as the Data Protection Act 1998. They also include formal governance through relevant policies, including Privacy Protection and Information Security policy; Safe Users of Research data Environments (SURE) Training Policy; terms of use which specify that the data must be used for research purposes only and the researcher may make no attempt to re-identify any individuals within the data[75]; and a policy which sets out the penalty for any breach of the terms of use.[76]

Although the purpose of these measures is in part to maintain legal anonymity, they would remain pertinent for administrative data research even if a different approach is taken to pseudonymised data. If, contrary to the arguments put forward in this article, it is determined at the UK or European level that all pseudonymised data are the data of identifiable individuals, and cannot be anonymised, good information governance will still be vital to ensure that those data are used legally, ethically and responsibly, and particularly to avoid direct identification of individuals. The importance of information governance is thus not an issue which stands or falls entirely on the GDPR definition of pseudonymised data. However, if anonymisation of pseudonymised data remains a possibility in law, it will help to reinforce good practice by defining identifiability with reference beyond original identifiers, to all the risks which good governance can address.

4. Conclusion

In conclusion, the GDPR should not be seen as expanding or re-defining the scope of personal data through its definition of pseudonymisation. The definition of pseudonymisation is not intended to be used to establish whether data are personal under the GDPR; indeed, it is clear that data to which pseudonymisation is applied are, and remain, personal data. Instead, it is Recital 26 GDPR which must be used to establish whether data are personal. This ultimately poses the question as to whether there exists a means reasonably likely to be used to identify natural persons. As such, anonymisation processes under the GDPR do not necessarily exclude pseudonymisation in the conventional sense, such as key-coding, as long as other environmental controls are in place to prevent the 'data situation' yielding identifiable data.

[71] Functional anonymisation asserts that one cannot determine the status of data as personal or not without reference to their environment, see the ADF at note 5.

[72] For further information on this please see the ADRN website at https://adrn.ac.uk/get-data/secure-access/, accessed 19 September 2017.

[73] <https://adrn.ac.uk/understand-data/sure-training> accessed 19 September 2017.

[74] Ibid.

[75] ADRN Terms of Use <https://adrn.ac.uk/media/174434/adrn021-termsofuse_v00-11_pub.pdf> accessed 19 September 2017.

[76] ADRN Breaches Policy <https://adrn.ac.uk/media/1297/adrn003_breachespolicy_02_pub.pdf> accessed 25 September 2017.

The case of *Breyer* even leaves the door open for data which have undergone pseudonymisation within one organisation, to be rendered anonymous from the perspective of an individual outside that organisation. This judgment demonstrates that the relationship between two parties is key to determining whether identifying information is sufficiently accessible for the data they hold to be deemed personal. This distinction is vital in a research context, as it leaves open the option of sharing 'anonymised' data with researchers, even when data are personal and GDPR pseudonymised within the organisation itself. Close attention to the terms of data sharing agreements will therefore be essential in a research context, in order to share data which would be confidential if identifiable, and to maintain the flow of data for public interest research in future.

Acknowledgements

The authors would like to thank Dr Mark Taylor, and two anonymous reviewers, for their comments on this article.

The authors received support from the ESRC grant number ES/L007452/1 for the Administrative Data Service. The funders played no role in the writing of this article, or the decision to submit it for publication.