COMPUTER LAW & SECURITY REVIEW ■■ (2018) ■■–■■

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.

Please cite this article in press as: Miranda Mourby, et al., Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK, Computer Law & Security Review: The International Journal of Technology Law and Practice (2018), doi: 10.1016/j.clsr.2018.01.002