Anonymous data v. Persona data—A false debate:

An EU perspective on anonymisation, pseudonymisation and personal data

Sophie Stalla-Bourdillon and Alison Knight

This era of big data analytics promises many things. In particular, it offers opportunities to extract hidden value from unstructured raw datasets through novel reuse. The reuse of personal data is however a key concern for data protection law as it involves processing for purposes beyond those that justified its original collection, at odds with the principle of purpose limitation.

The issue becomes balancing the private interests of individuals and realising the promise of big data. One way to resolve this issue is to transform the personal data to be shared for further processing, *e.g.* data mining, into "anonymous information," to use an EU legal term. "Anonymous information" is outside the scope of data protection laws in the EU, and is also carved out from privacy laws in many other jurisdictions worldwide.

The foregoing solution works well however as long as the output potential from the data still retains utility, which is not necessarily the case. This is because the value or knowledge that can be gained from analysing datasets using software is maximised by virtue of finding patterns, basically linking relationships between data points. Anonymisation, by contrast, aims to delink such data point relationships where they relate to informational knowledge that can be gleaned in respect of specific persons and their identities. This leaves those in charge with the processing of the data with a problem: how to ensure that anonymisation is conducted effectively on the data on their possession, while retaining that data's utility for potential future disclosure to, and further processing by, a third party.

Despite a broad consensus around the need for effective anonymisation techniques, the debate as to when data can be said to be legally anonymised to satisfy EU data protection laws is a long-standing one. Part of the complexity in reaching consensus on this issue derives from confusion around terminology, in particular the meaning of the concept of anonymisation in this context, and how strictly delineated that concept should be. This is, in turn, due to the lack of consensus about the doctrinal theory that underpins its traditional conceptualisation as a privacy-protecting mechanism.

For example, the texts of both the existing EU Data Protection Directive[1] (DPD) and the new EU General Data Protection Regulation[2] (GDPR) are ambiguous. While both legal instruments seem to adopt a restrictive definition of "anonymous information", anonymous in such a way that the data subject is no longer identifiable as it will be described below, they also seem to support a risk-based approach that limits the identifiable concept by a reasonableness standard, i.e. to the extent that *only* all the means *likely reasonably* to be used to identify someone are taken into account. [3]

In fact, the GDPR seems even more restrictive than the DPD, for a new category of data (i.e. data that has undergone pseudonymisation)–to be distinguished from the category of "anonymous information", or data that has undergone anonymisation to use a better terminology[4]–is introduced.[5] The concept of pseudonymisation is defined under the GDPR as meaning:

> *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*

[1] Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data.

[2] The GDPR was agreed by the European Commission, European Parliament, and the Council of the EU in December 2015 to replace the DPD. In April 2016, the European Parliament formally approved the final text version of the GDPR for translation into the EU's official languages. The GDPR – formally entitled, 'Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC' - was subsequently published in the EU Official Journal on 4 May 2016 (OJ L 119, 4.5.2016, p. 1, available at http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL%3A2016%3A119%3ATOC, retrieved on 17 May 2016). It comes into force on 24 May 2016 and takes effect on 25 May 2018. Thus, organisations have two years within which to ensure that they comply with the GDPR in anticipation of that date.

[3] Recital 26 of the DPD, as well as Recital 26 of the GDPR.

[4] We will, however, use the term "anonymised data" in this article as it is shorter and makes the reading easier. We are nonetheless of the view that the expression "data that has undergone anonymisation" better captures the idea of data characteristics as fluid concepts which, as a matter of fact, can only be understood in the context of appreciating ongoing processes related to the data environment, and which does not 'simply' focus upon data as having static and immovable qualities as we will explain below. A similar choice has been made by others. See e.g. M. Elliot, E. Mackey, K. O'Hara and C. Tudor, The Anonymisation Decision-Making Framework, 2016, University of Manchester: Manchester.

[5] Recital 26 of the GDPR.

> *and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*.[6]

Personal data that has undergone pseudonymisation is explicitly stated to remain personal data under EU data protection laws as it "*should be considered to be information on an identifiable natural person*."[7]

It is suggested in this paper that while the concept of anonymisation is crucial to delineate the scope of data protection laws at least from a descriptive standpoint, recent attempts to clarify the terms of the dichotomy between "anonymous information" and personal data (in particular, by data protection regulators in the EU) have partly failed. While this failure could be attributed to the very use of a terminology creating the illusion of a definitive and permanent contour that would clearly delineate the scope of data protection law, the reasons of such a failure are slightly more complex. Essentially, it can be explained by the implicit adoption of a static approach, which tends to assume that once the data is anonymised, not only can the initial data controller forget about it, but also the recipients of the dataset are free from any obligation or duties since the transformed dataset, once and for all, lies outside the scope of data protection laws. By contrast, the state of anonymised data has to be comprehended in context, which includes an assessment of: the data; the infrastructure; and the agents[8]. Moreover, it is very important to comprehend the state of anonymised data dynamically. This dynamic state is epitomised by the fact that anonymised data can become

[6] Article 4(5) of the GDPR.

[7] To note, while the final GDPR text does not make pseudonymous data (so defined) a special category of personal data - in the sense that it does not seem to benefit from a light-touch data protection regime in being exempted from certain data protection rules, doubts still persist over the exact implications of its status. For example, see the formulation of Recital 29 of the GDPR, which states, "[i]*n order to create incentives for applying pseudonymisation when processing personal data, measures of pseudonymisation whilst allowing general analysis should be possible within the same controller when the controller has taken technical and organisational measures necessary to ensure, for the respective processing, that the provisions of this Regulation are implemented, and ensuring that additional information for attributing the personal data to a specific data subject is kept separately*". Besides Article 6(4) provides that in order to ascertain whether further processing is compatible with the purpose of the initial processing, considerations relating to the existence of appropriate safeguards such as encryption or pseudonymisation should be taken into account.

[8] M. Elliot, E. Mackey, K. O'Hara and C. Tudor, The Anonymisation Decision-Making Framework, 2016, University of Manchester: Manchester, p. 2. *"The framework is underpinned by a relatively new way of thinking about the re-identification problem which posits that you must look at both the data and the data environment to ascertain realistic measures of risk"*.

personal data again, depending upon the purpose of the further processing and future data linkages.

The implications of this dynamic approach are, in particular, that recipients of anonymised data, while 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. In particular, the former must abide by any licensing limitations upon the purpose and the means of the processing of such data in its disclosed post-anonymisation process form to remain outside the scope of data protection laws. At the same time, the very characterisation of anonymised data should also be dependent upon an ongoing monitoring on the part of the initial data controller of the data environment of the dataset that has undergone anonymisation.

The paper starts by examining the recent approaches to anonymisation, highlighting in particular the shortcomings of the legal and technical approaches to this issue adopted at the EU level, which are based implicitly on a static approach, assuming that once the anonymised dataset is released the recipient has complete freedom of use over its subsequent processing. The paper then unfolds the main component of a dynamic approach, and explains why an approach to anonymisation of this type is both more appropriate and compatible with the current and soon-to-be-applied EU legal framework under the GDPR. Ultimately, the paper makes the point that the opposition between so-called “anonymous information” and personal data in a legal sense is less radical than usually described.