European Journal of Law and Technology, Vol 10, Issue 1, 2019

EJLT European Journal of Law and Technology

Rethinking the "release and forget" ethos of the Freedom of Information Act 2000: Why developments in the field of anonymisation necessitate the development of a new approach to disclosing data

Henry Pearce[1] and Sophie Stalla-Bourdillon[2]

Abstract

The Freedom of Information Act 2000 (FOIA) gives individuals the right to request and receive access to information held by public authorities. Under the FOIA, a public authority releasing requested information has no post-release obligations to monitor any subsequent uses of that information, nor are any specific obligations imposed on the recipient of the information. It is made clear in the FOIA, however, that in most circumstances any information that constitutes personal data (i.e. any information relating to an identified or identifiable living individual) will be exempt from freedom of information requests.

In the last few years the interplay between freedom of information requests and data protection law has been considered by UK courts in several interesting cases. By and large, these cases have focused on issues relating to the anonymisation of personal data. Under UK and EU data protection legislation data that have been anonymised so that they can no longer be used to identify an individual are considered anonymous, and thus not personal data. As anonymous data are not personal data they are not exempt from freedom of information requests made under the FOIA. Operating under this premise, UK courts have begun to order public authorities to release datasets containing anonymised personal data to individuals who have requested access.

As the FOIA imposes no post-release obligations on the releaser or recipient of requested information it can be said to endorse a "release and forget" approach to disclosing data. In the context of datasets containing anonymised personal data, however, this approach is problematic. Recent work undertaken in the field of anonymisation has revealed that total and infallible anonymisation of personal data is not possible. Instead, it has been convincingly demonstrated that anonymisation is highly context-dependant, and that the success of attempts to anonymise data will be contingent on a range of factors such as the environment into which the data are to be released, how that environment might change over time, the identity and range of the recipients of the data, and the future purposes to which those data will be turned. As a result, the "release and forget" approach upon which the FOIA appears to be premised is not fit for purpose.

The function of this article is twofold. First, it argues that the approach to anonymisation and personal data taken by the FOIA is detached from contemporary authoritative understandings of these concepts and should be rethought. Second, having outlined the limitations of the current approach, the article proposes a new model for disclosing data under the FOIA based on notions of privacy and data protection by design.

Introduction

The Freedom of Information Act 2000[3] gives individuals the right to request and receive access to information held by public authorities. Under the FOIA, a public authority releasing requested information has no post-release obligations to monitor any subsequent uses of that information, nor are any specific obligations imposed on the recipient of the information regarding its future uses or further sharing. As such, the FOIA can be said to embrace a "release and forget" approach to disclosing data. It is made clear in the FOIA, however, that in most circumstances, information that constitutes personal data as per the definition used by UK data protection legislation is exempt from FOI requests. Conversely, personal data that have been rendered anonymous (i.e. "non-personal") through a process of anonymisation are not exempt from FOI requests.

However, in recent years questions regarding whether, and to what extent, the anonymisation of personal data is possible have become increasingly prominent. Recent advances in information technologies and analytical tools have demonstrated that the complete and infallible anonymisation of personal data is not possible.[4] Predictably, this has caused difficulties for data controllers and public authorities when it comes to determining whether data at their disposal can be considered "personal" or "anonymous". Significantly, however, both EU and UK data protection legislation appears to allow for a risk-based approach to the concept of personal data, according to which data will only be considered "personal" if there is a significant risk of those data being used to identify an individual.[5] As a result, attempts to determine whether data can be considered "personal" or "anonymous" have become exercises in risk management.

Though questions surrounding anonymisation and its shortcomings have been discussed in academic and scholarly literature for the best part of a decade,[6] they have recently assumed greater practical significance. UK courts, for instance, have increasingly been asked to consider how these exercises in risk-management can and should be undertaken in practice, particularly in relation to FOI requests. Specifically, courts are increasingly being asked to consider whether data at the centre of FOI requests can be considered anonymous, and thus suitable for disclosure, or personal, and therefore exempt from FOI requests. For instance, in one notable case, *Queen Mary v Alem Matthees*,[7] the First Tier Tribunal[8] held that data in the possession of Queen Mary, University of London, were anonymous and subsequently ordered that they be disclosed pursuant of the applicant's FOI request.

By way of reference to the UK Anonymisation Network's[9] Anonymisation Decision-making Framework,[10] and other recent research in the field,[11] this article argues that the approach of the FTT in this case, whilst consistent with the core tenets and provisions of the FOIA, was based on an approach to anonymisation that was disconnected from leading authoritative understandings of the concept. This, the article contends, highlights a range of problems inherent in the FOIA's "release and forget" ethos, and that this approach to data disclosure needs to be fundamentally rethought. Following this, the article proposes and outlines a new alternative

model for disclosing data under the FOIA which has been designed specifically to address some of the most notable shortcomings of the current approach.

The article takes the following structure. First, the article outlines the legislative background to the abovementioned issues and explains the FOIA's right of access, the concept of personal data, and the interplay between the two. Second, the article explains the concept of anonymisation, how total and infallible anonymisation of personal data is not possible, but how UK and EU data protection law appears to allow for a risk-based approach to the categorisation of personal and anonymous data. Third, the article outlines and explains the facts, issues and judgment of *Queen Mary v Alem Matthees*. Having done this, the fourth section of the article then argues that, in light of recent developments in the field of anonymisation, not only must the approach taken by the FTT in this case be considered incorrect, but that it serves as a timely demonstration of how the FOIA's release and forget approach to data disclosures itself is unfit for purpose. In its final substantive section, the article posits how the limits of the FOIA's "release and forget" disclosure model might conceivably be addressed by way of the adoption of data protection by design strategies. In this section, the article proposes a new model for disclosing data under the FOIA that incorporates a range of data protection by design elements. The article concludes with a summary of its core arguments.

1. The Freedom of Information Act 2000

Having reached the statute book in 2000 the FOIA came into force in January 2005,[12] helping to crystallise various provisions of the European Convention on Human Rights,[13] notably the right to freedom of expression,[14] into UK law. As has been noted elsewhere, perhaps the most significant driver behind the FOIA's enactment was a desire to give effect to the principle of open and transparent government.[15] The most noteworthy way in which the FOIA is geared towards this objective is the way in which it bestows upon individuals a general right of access to information held by public authorities. This is specified in section 1(1), which states:

> "*Any person making a request for information to a public authority is entitled—*
>
> *(a) To be informed in writing by the public authority whether it holds information of the description specified in the request, and*
>
> *(b) If that is the case, to have the information communicated to him*."

Pursuant of this, presumably to ensure maximum coverage and avoid the emergence of loopholes, the FOIA gives a very wide definition to the term "public authorities". In contrast with other pieces of UK legislation, which tend to treat the term with ambiguity,[16] the FOIA includes a specific list of all institutions, actors and bodies that can be considered a public authority. This list includes, but is not limited to, central government departments and agencies; local authorities; National Health Service bodies, including individual general practitioners, dentists and pharmacists; schools, colleges and universities; the police and armed forces; regulators; publicly owned companies; and the British Broadcasting Corporation.[17]

Upon receiving a FOI request as per section 1(1), assuming no exemption applies, a public authority must identify and disclose the requested information to the requesting party within twenty days of the date of request.[18] A failure to do so will be unlawful. The recipient of the information released by the public authority (i.e. the party responsible for making the FOI request) will then be under no obligations so far as future re-uses, sharing and disclosures of that

information are concerned. In other words, so long as no other statutory or common law restrictions apply, the recipient of the information is free to use that information in any way they wish, including making that information open to the public.[19] Concurrently, the public authority is under no obligation to monitor any uses of that information post release. To this end, it can be said the FOIA embraces a "release and forget"[20] approach to data disclosures. The implications of this will be returned to later in the article.[21]

What is significant for our present purposes are some deliberate caveats to the FOIA's right to information. Specifically, Part II of the FOIA adopts a wide range of exemptions which exclude various types of information from requests made under section 1(1).[22] The exemption that is of greatest interest to this article, however, is information that is personal data.[23]

1.1. The interplay between the FOIA and data protection law

The now superseded Data Protection Act 1998[24] referred to in the FOIA defined personal data as:

> "*…data which relate to a living individual who can be identified—*
>
> *(a) from those data, or*
>
> (b) *from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller…*"[25]

Section 3 of the recently enacted Data Protection Act 2018,[26] which replaces the 1998 Act, offers a more comprehensive definition which mirrors that contained within the General Data Protection Regulation[27] of the European Union:

> "*(2) "Personal data" means any information relating to an identified or identifiable living individual…*
>
> *(3) "Identifiable living individual" means a living individual who can be identified, directly or indirectly, in particular by reference to—*
>
> *(a) an identifier such as a name, an identification number, location data or an online identifier, or*
>
> *(b) one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of the individual*."[28]

Section 40(2) of the FOIA then specifies that personal data is exempt from any requests made under section 1(1) FOIA if disclosure of those data would contravene any of the data protection principles as contained within the DPA 2018. Specifically, the principles hold that: the processing, including the disclosure, of personal data must be lawful, fair and transparent;[29] that the purpose for which personal data is collected must be specified, explicit and legitimate, and that personal data must not be processed in any way that is incompatible with any such purpose;[30] personal data must be adequate, relevant, and not excessive in relation to the purpose for which they are processed;[31] personal data must be accurate and, where necessary, kept up to date;[32] personal data must be kept for no longer than is necessary for the purpose for which it is processed;[33] and personal data must be processed in a manner that includes taking appropriate security measures as regards risks that arise from processing personal data.[34]

The principle that is most obviously applicable in the immediate context is that which states that the processing of personal data must be lawful and fair. For any processing of personal data to be considered lawful, one of the legitimising grounds mentioned in Article 6(1) GDPR must apply:

(a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

(b) Processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

(c) Processing is necessary for compliance with a legal obligation to which the controller is subject;

(d) Processing is necessary in order to protect the vital interests of the data subject or of another natural person;

(e) Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

(f) Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data… [35]

As has been noted elsewhere, to date when dealing with section 40(2) FOIA cases, in determining whether it will be lawful to disclose personal data as part of an FOI request public authorities have tended to invoke the 'legitimate interests' ground for processing as their relevant justification as per Schedule 2 of the DPA 1998.[36] Whilst Article 6(1)(f) of the GDPR contains a restriction which prohibits public authorities from invoking the "legitimate interests" grounds for processing "in the performance of their tasks",[37] the new section 40(8) of the FOIA inserted by the DPA 2018 disqualifies this restriction in the context of disclosing data following FOI requests.[38]

So far as the fairness of any processing of personal data is concerned, the Information Commissioner's Office, the UK's independent regulatory body concerned with overseeing matters of data protection and electronic communications,[39] has published various guidance notes on how fairness can be assessed.[40] In its guidance on requests for personal data of public authority employees, for instance, the ICO advises that a range of factors could indicate whether the disclosure of personal data in the context of a FOI request would be fair. Specifically, it is suggested that if information at the centre of an FOI request is sensitive personal data its disclosure is unlikely to be fair.[41] Other factors, such as the reasonable expectations of the individuals to whom the data relate, the rights and freedoms of those individuals, and the weight of possible legitimate interests in the disclosure of the data, are also highly relevant when it comes to determining fairness in this context.[42] In short, what this amounts to is the fact that, due to circumstances surrounding data at the heart of an FOI request, many FOI requests regarding personal data, including those that are based on the "legitimate interests" of other parties, may be outweighed and defeated by the interests of those to whom the data relate, meaning that disclosing personal data under the FOIA will often be difficult to achieve.

In any event, when read in conjunction with one another, the definition of personal data contained in the DPA 2018, section 40(2) FOIA, the relevant parts of the GDPR, and the advice of the ICO, have a clear combined effect: if information held by a public authority can be used to identify a living individual (i.e. the information is personal data), and the disclosure of this

information either does not fall within one of the GDPR's legitimising grounds or would be unfair to the individuals to whom it relates, it will be exempt from requests made under section 1(1) FOIA. The only way data of this sort can be disclosed under the FOIA is if they are anonymised, so that they are no longer "personal". Since the coming into force of the FOIA in 2005, however, this has proved to be a source of contention.

On several notable occasions, for instance, public authorities have refused to grant requests for information made under section 1(1) FOIA on the basis that the requested information amounted to personal data, and its disclosure would breach the data protection principles, leading to them being brought before the courts due to alleged misapplications of s.40(2).[43] Most of these disputes have revolved around disagreements as to the nature of the information requested, with the requesting party alleging that the information sought does *not* constitute personal data, the public authority of whom the request has been made arguing that it *does* and that disclosure of the data would be unlawful and/or unfair, and the courts being asked to decide one way or the other.

At a glance this might not seem especially troublesome. It might be assumed, for instance, that determining whether information can or cannot be used to identify an individual, or whether information "relates" to an individual, should ordinarily be straightforward. This, however, is often far from the truth. Due to constant developments in the field of information technology that have occurred since the mid-twentieth century, 'personal data' is a term capable of encompassing an ever-expanding range of information. As a result, discerning when and whether certain types of information constitute personal data can be extremely challenging. This is particularly the case in situations where personal data have supposedly been anonymised.

2. Anonymous data and the shortcomings of "anonymisation" techniques

As explained above, the exemption listed under section 40(2) FOIA applies to data that are "personal" in accordance with the definition used by data protection legislation. However, as is made explicitly clear in the recitals of the GDPR anonymous data should not be considered personal data:

"*The principles of data protection should…not apply to anonymous information, namely information that does not relate to an identified or identifiable natural person or to data rendered anonymous in such a manner that the data subject is no longer identifiable*."[44]

Ergo, data that are non-personal, or have been rendered anonymous by way of anonymisation techniques do not fall within section 40(2)'s remit, and thus are not exempt from FOI requests.[45] As is considered in greater detail below, for data to be considered anonymous the risk of an individual being identified from those data must be negligible.

In an ideal world this would not give us much reason for pause. If it were possible, for instance, to definitively sort every datum into boxes marked "personal" and "anonymous" there would be no complications inherent in the law's dichotomous treatment of these two data types. Unfortunately, however, this ideal is far-removed from reality. Notably, due to advances in so-called "re-identification techniques" that have occurred over the last decade or so, a development which can primarily be attributed to technological developments such as the emergence of "big data",[46] the idea that there is such a thing as complete, infallible and perfect anonymisation of

data has been exposed as a myth.[47] Put simply, it is no longer possible to guarantee the absolute anonymisation of data through the use of traditional anonymisation techniques over time, as the combination and cross-analysis of supposedly anonymised data with other data, which may or may not be publicly available, will in some cases make it possible for individuals to be "re-identified" from data that have supposedly been anonymised.[48]

Helpfully, the Article 29 Working Party's Opinion on Anonymisation Techniques[49] describes three common types of re-identification risks:

'**Singling out'**: the possibility of isolating some, or all, of the records contained in a dataset which identify an individual within that dataset.

'**Linkability'**: the ability to link at least two records containing the same data subject or a group of data subjects, either in the same database or in two different databases.

'**Inference'**: the possibility of deducing, with significant probability the value of certain attributes in a dataset from the value of other attributes.[50]

As is explained below, the extent to which any dataset or data are susceptible to these risks will depend on context and a range of different factors. In any event, "anonymous" data as interpreted by the Article 29 Working Party are data for which all three of the abovementioned types of risk are mitigated.

As has been noted elsewhere, however, the emergence of these risks has blurred the boundaries of the dichotomous approach to personal and anonymised data taken by European and UK data protection legislation.[51] This has also led to proclamations that the practical worth of anonymisation techniques is now effectively zero and, as a result, the abovementioned dichotomy between personal and anonymised data should be abandoned, as the dividing line between "personal" and "anonymous" data has become irreversibly blurred.[52] Others, however, have been more optimistic and have highlighted how whilst complete and infallible anonymisation of data is not possible, successful anonymisation can still be achieved in some situations.[53]

Pursuant of this, various observers, having highlighted that EU data protection law does not require a one hundred percent guarantee of non-identifiability in order for data to avoid being categorised as personal data,[54] have advocated for the adoption of a risk-based approach to data protection, and argued that data should only be considered personal if there is a substantial risk of those data being used to identify an individual.[55] Put another way, under this approach data should be considered "anonymous" if the risk of them being used to identify an individual is negligible. As has been noted elsewhere, a close analysis of the provisions of the GDPR reveals that it appears to allow for such an approach.[56]

Consequently, regulatory bodies and other organisations have begun to offer advice on how such determinations can be made, and how anonymisation should be understood. Notably, in 2012 the ICO published its code of practice on the anonymisation of personal data.[57] The code contains guidance on how data controllers can take steps to minimise the risk or re-identification associated with personal data they have anonymised, and states that if, having considered a range of relevant factors, the risk of re-identification associated with data is remote or negligible, then those data should be considered anonymous, not personal.

Of note, the code suggests that data controllers deploy a "motivated intruder" test, to see whether a motivated intruder with no specialist knowledge or equipment would likely be able to

identify an individual from data without having to resort to criminality.[58] If the answer to this question is "no" this, according to the ICO, should indicate that the data are anonymous. Whilst not binding in law, the code is generally thought to be persuasive, and has come to be used by the courts in cases relating to FOI requests as a means of determining whether data at the centre of disputes between public authorities and parties responsible for making FOI requests can be considered anonymous or personal. However, perhaps the most authoritative guidance issued on anonymisation to date, and certainly that which is most relevant to this article, is the Anonymisation Decision-making Framework of the UK Anonymisation Network.[59]

2.1. The UKAN Anonymisation Decision-making Framework[60]

The UKAN Framework represents the culmination of a three-year cross-sector collaboration process between multiple disciplines. Though its 156 pages are not legally binding, their contents and guidance are widely considered to be authoritative.[61]

Whilst noting that anonymisation is generally not a well-understood concept, the Framework explains that it is an ongoing area of research that is of critical importance, and that whilst it faces many complex issues and challenges, the majority of these should not be thought of as being insurmountable. More generally, whilst acknowledging that complete and infallible anonymisation of data is not possible, the Framework reiterates the argument that anonymisation is still achievable in some situations and contexts. In so doing, it provides a guide for those in possession of datasets which include individuals' personal data in respect of how to anonymise these data, reduce and balance risks of re-identification, and develop best practice.

The Framework's key overarching message is that anonymisation is a heavily context-dependant process, and that it is only by considering anonymised data *and* the environment into which they might be released that a decision can conscientiously be made as to whether said data can be considered anonymous. The contextual details surrounding the release of anonymised dataset are referred to by UKAN as "the data situation".[62] Historically many approaches to anonymisation have tended to rely on looking at data alone, in isolation, ignoring the data situation of those data. Such approaches, according to UKAN, are premised on an outdated understanding of anonymisation and should be resisted.[63]

So to justify this position, the Framework notes that it will be impossible for holders of anonymised datasets to guard against risks of re-identification unless they have some knowledge in respect of the nature of those risks, an appreciation of their severity, and the likelihood of them manifesting.[64] A consideration of the nature of the data, the recipients of the data, the purposes for which the data will be used, the existence and incentives of "motivated intruders", and the existence of other types of data available either publicly or privately both in the present and in the future, will therefore be critical to determining whether data can be considered anonymous. To this extent, the Framework's guidance can be said to align somewhat with the abovementioned ICO anonymisation code of practice.

Perhaps the most significant aspect of the UKAN Framework, however, is the way in which it convincingly argues and strongly emphasises that anonymisation should be thought of as an ongoing process, or function of data and their external environment, rather than as a one-time procedure or finite end state.[65] In other words, attempts to anonymise data should be thought of as continuously enduring activities, that should not necessarily ever be thought of as being

definitively concluded. In this sense, the Framework explains how the probability of re-identification of certain data might conceivably be considered minimal at the time of their release (i.e. the data could be considered anonymous), but at a later point in time changes to the "data situation" brought about by technological developments or the availability of new data sources could, post-release, increase the risk of those data being re-identified to a higher level, effectively rendering them "personal" once more.[66]

The Framework suggests, therefore, that the releasers of anonymised datasets should continue to observe the environment into which such datasets are released to monitor whether the risks of re-identification remain low and, if the level of risk rises, take steps to mitigate those risks post-release. It is only through taking such steps that successful anonymisation can truly be achieved.[67] According to UKAN, therefore, the issue of data identifiability (i.e. whether data are personal or anonymous) is relative to the contextual situation of those data at any given time. This approach is referred to by UKAN as "functional anonymisation".[68] As has been noted elsewhere, due to its holistic ethos, UKAN's approach can also be described as "dynamic".[69]

This is an approach to the concept of personal data that appears to be consistent with recent jurisprudence of the Court of Justice of the European Union.[70] This is best shown by the recent *Breyer*[71] case, where the CJEU concluded that Internet Protocol addresses would only constitute personal data in the hands of a website operator if another party, such as an Internet service provider could link an address to an individual, and if the website operator had a legal means of obtaining the information held by the service provider.

Troublingly, however, despite UKAN's authoritative guidance, there are reasons to believe that anonymisation is poorly understood by the courts, and moreover, that the FOIA's "release and forget" approach cannot be considered appropriate in a world where anonymisation should be thought of as a dynamic and highly context-dependant process. The case of *Queen Mary v Alem Matthees* serves as a prime example as to why this is so.

3. *Queen Mary University of London v (1) The Information Commissioner and (2) Alem Matthees*, EA/2015/0269

3.1. Facts

The case revolved around a long-term and large-scale clinical research trial conducted by Queen Mary, University of London,[72] informally known as "PACE".[73] The trial, which commenced in 2002, was designed to test the effectiveness of various treatments for the sufferers of Chronic Fatigue Syndrome. As part of the trial many medical baseline and treatment results were collected from over six hundred trial participants, with much of this data being collected via questionnaires. The data gathered did not contain any direct or indirect identifiers (e.g. data regarding participants' names, location, gender or ethnicity), and participants were assured that the confidentiality of those data would be guaranteed. The participants gave their consent, however, to those data potentially being anonymised and shared with independent scientists from outside the university upon request for the purposes of collaborative research, subject to additional confidentiality agreements.[74]

The results of the trial were controversial and its methodology was widely criticised.[75] Pursuant

of this, the complainant, Mr Matthees, made a request under section 1(1) FOIA to access anonymised aspects of the data used in the trial with the intention of testing the PACE trial methodology. The request was rejected by QMUL for several reasons. The reason that is of the greatest interest to this article, however, was the fact that the data in question related to individuals who had participated in the trial and, as a result, constituted "sensitive personal data" as per the definition contained within the Data Protection Act 1998. As there was no legitimising ground for disclosing such data, and that doing so would be unfair to the individuals to whom the data pertained (i.e. disclosure would breach the data protection principles), QMUL declared the PACE data immune from disclosure on the basis of s.40(2) FOIA.

After a process of internal review, QMUL's decision to reject Mr Matthees' request was appealed to the UK Information Commissioner.[76] Following an examination of the facts, and the different possible risks of re-identification, the IC concluded that the data to which Mr Matthees' access request pertained had successfully been anonymised and thus were not sensitive personal data for the purposes of the DPA. Therefore, the PACE data were not covered by s.40(2) FOIA.[77]

Specifically, the IC pointed out that the PACE dataset contained one row for each of the 640 participants, each row had fourteen columns, one of which had a PIN number representing a single trial participant, with the others representing numerical scores for the various outcomes of the tests undertaken. Though the IC conceded that in theory it would be possible for individuals to be re-identified from this information were it to be released publicly (e.g. participants of the trial, or possibly people known to them, could re-identify by way of locating their own scores in the dataset, or re-identification occurring through the PACE data being linked with other medical records) this was very unlikely to occur, and that in order for the data to be considered personal the risk of re-identification must be greater than remote.[78]

Accordingly, the IC concluded that QMUL's argument that the PACE data constituted sensitive personal data was based on an error of law. As there was no lawful reason for refusing Mr Matthees' FOI request, the IC ordered that the data must be disclosed. QMUL appealed to the FTT.

3.2. Issues

On appeal, the FTT was asked to resolve whether the data requested by Mr Matthees were personal data for the purposes of the DPA 1998, and whether even though those data had formally been anonymised, there was a chance that participants in the trial could be identified from the requested data.

In relation to these questions, QMUL argued that despite efforts having been made to anonymise the data, there was a significant risk that a "motivated intruder" would be able to re-identify individuals from the data were they to be released openly and without restriction. This, it was argued, was enough to demonstrate that the PACE data were indeed sensitive personal data, and thus, that they were exempt from disclosure under the FOIA.

3.3. The First Tier Tribunal's Decision

By a two to one majority, the FTT rejected QMUL's argument that the data were sensitive personal data, and instead agreed with the IC's conclusion that the data had been anonymised and should be disclosed. In making its decision the FTT endorsed the risk-based approach to the concepts of anonymisation and personal data advanced by the ICO in its abovementioned code of practice,

and held that where personal data have been anonymised, so long as the risk of re-identification is "remote" those data should not be considered "personal". Operating under this premise the FTT held that the risk of individuals being identified from the PACE data, were those data to be released publicly, was remote. Ergo, the PACE data could rightly be considered anonymous and not personal. This conclusion was based on several factors.

First, the FTT pointed out that the data in question had been subject to anonymisation techniques and contained no direct identifiers. Instead, the data were based on variable outcomes that would be difficult to repeat with any great precision (e.g. various physical tests).[79] This, it was argued, was suggestive of the re-identification risk attached to those data being low.

Second, the FTT observed that the possibility of third parties re-identifying PACE participants from an analysis of the anonymised data in question *alone* was essentially nil. The only way for re-identification to occur would be for a third party to discover other information that could be used to link the anonymised records contained in the PACE dataset with records in another dataset, the likes of which would not, in theory, be available to the public. In relation to this finding, the FTT concluded that for the data within the trial dataset to be linked to a trial participant the expertise and knowledge of a medical professional who was willing to breach their professional and legal obligations would most likely be required. Noting that according to the ICO code of practice on anonymisation suggests that a "motivated intruder" should not presumed to have any "specialist knowledge or equipment", nor should they be "expected to resort to criminality to access the data",[80] the FTT suggested it was implausible that the data would ever end up in the hands of such a person, and rejected QMUL's argument that a motivated intruder could possibly re-identify individuals from the data.[81]

It was also remarked that when attempting to determine the risk of re-identification associated with an anonymised dataset it was not necessary to take into account efforts that were "borderline sociopathic or psychopathic" in nature.[82] In relation to this point, the FTT was also keen to stress that whilst external factors that may impact on the risk of re-identification must be considered, such as the existence of motivated intruders and availability of other information, generic references to social media and any other non-specific assertions (e.g. because there are so many data "out there" etc.) are not sufficient to show that a risk of re-identification is more than "remote".

Third, the FTT noted that historically, other similarly anonymised clinical trial datasets had been released publicly without giving rise to any significant re-identification incidents. This, it was argued, was suggestive of the disclosure of the PACE data also being low risk.[83]

Finally, as QMUL had already shared the PACE data with independent scientists for the purposes of collaborative research, it was argued that this amounted to an implicit acknowledgement on behalf of QMUL that the PACE data had been successfully anonymised, thereby indicating the suitability of disclosing those data more widely.

4. Dissecting the decision of the FTT in the QMUL case

There are elements of the FTT's decision that give rise to some substantial concerns. These primarily relate to the way in which aspects of the FTT's reasoning was disconnected from contemporary authoritative understandings of anonymisation. Specifically, this section of the article critiques the FTT's decision on the basis that:

- The FTT deployed a simplistic and outdated version of the ICO's motivated intruder test;
- The FTT adopted an overly simplistic view of the importance of the external environment into which the PACE data were to be released which failed to account for the whole data situation of the PACE data; and
- The FTT drew improper conclusions about the risk-status of the PACE data from other unrelated data disclosures.

Before proceeding further, it is important to note that this section of the article is not designed to solely criticise the judgment of the FTT, but to show how the questionable approach of the FTT in the QMUL case exposes the unsatisfactory nature of the FOIA's "release and forget" approach. Having concluded that the requested PACE data had been satisfactorily anonymised, after all, the FTT had no option but to order the unfettered, "release and forget", disclosure of those data due to the wording of the substantive provisions of the FOIA. As is explained below, however, the unrestricted and completely open release of anonymised datasets cannot be considered good practice in most circumstances. In other words, though the FTT's approach and reasoning in and of itself was dubious, its decision illustrates the bigger point of how the legislative framework under which it was operating is unfit for purpose and in need of reform.

4.1. The deployment of a simplistic and outdated version of the ICO's motivated intruder test

As noted above, the ICO's default version of the motivated intruder test holds that a motivated intruder should not be taken to have any specialist knowledge, nor should they be expected to resort to any criminality to re-identify anonymised data. Pursuant of this, as the re-identification of the PACE data would likely require expert knowledge and/or criminal conduct, the FTT deemed that the data was sufficiently anonymised as a motivated intruder, as envisaged by the ICO code of practice at least, would not be able to re-identify any individuals from an analysis of those data.

As implied by the UKAN Framework, however, in some contexts the releaser of a dataset may have to adopt a different standard of the motivated intruder test depending on the nature and sensitivity of data they seek to release. Some data types for instance, will be much more attractive to certain parties than others, and thus may attract more highly motivated and/or skilled intruders. A consideration of the motivations of such persons will be critical to developing an assessment of the data environment of those data.[84]

Particularly, the Framework notes that the more sensitive the data in question are, the more likely it is that they will attract intruders with extensive skills and motivations, and hints that the motivated intruder test should be adjusted in accordance with this. Given the highly sensitive nature of medical and healthcare data, it seems plausible that the PACE dataset could conceivably have been more attractive to intruders, including those that might be willing to resort to criminality, than a dataset containing data of a lesser sensitivity. Accordingly, it is concerning that the FTT, and the IC before it,[85] were happy to rely on the default version of the motivated intruder test, and dismissed the idea of the nature of the data contained within the PACE dataset necessitating the adoption of a higher standard in accordance with the data situation.[86] The fact, for instance, that ICO code holds that a motivated intruder should not be expected to resort to commit a criminal offence in order to re-identify data raises questions about its general suitability, and in so doing the suitability of the approach of the ICO and FTT, as it seems far from implausible

that in instances involving sensitive and/or valuable data an intruder could resort to such means.

4.2. The adoption of an overly simplistic and limited view of the importance of the external environment into which the PACE data were to be released

Irrespective of the appropriateness of the way in which questions surrounding the motivated intruder test were dealt with, a further concern to be had with the judgment is the fact that the FTT appeared to treat the assessment of the re-identification risks associated with the PACE data as a "one and done" type exercise, rather than considering anonymisation an ongoing process. Essentially, in finding that the risk of re-identification associated with the data was remote, the FTT appeared to be of the belief that as the data could be considered anonymous at the time the case was heard, the data would remain anonymous for the remainder of their existence.

As noted above, in reaching this conclusion the FTT was critical of suggestions that vague and general statements about re-identification risk being high due to so many data being available "out there". This must surely be correct, as allowing for mere speculation to inform decisions about re-identification risk cannot be considered good practice.[87]

However, whilst the FTT rightly eschewed the overly simplistic approach of arguments made on such grounds, it failed to meaningfully consider possible future changes to the data situation of the PACE data which, as explained previously, are critical to determining whether data can be considered anonymised. As noted above, whilst the risk of re-identification of an anonymised dataset might rightly be categorised as low at the point of disclosure, changes to the data environment, such as the availability of new data or emergence of new technologies, could mean that in time the level of risk associated with that dataset could rise. The fact that the FTT did not consider this possibility in any meaningful way further highlights how its decision appeared to be made on an analysis and consideration of the data contained within the PACE dataset alone, with scant attention being paid to the external environment of those data. For the reasons outlined earlier in the article, this approach is plainly inconsistent with contemporary authoritative understandings of anonymisation.

4.3. The drawing of improper conclusions about the PACE data from other unrelated data disclosures

Similar observations can also be made in respect of the FTT's suggestion that the fact that other similarly anonymised unrelated datasets had been previously released without giving rise to re-identification problems was indicative of the PACE data also being of a low re-identification risk. However, this was neither necessarily here nor there so far as determining the level of risk associated with the PACE data was concerned, due to differences between the data environments of the immediate case and previous instances. As is convincingly argued by the UKAN Framework, for anonymisation attempts to have any chance of succeeding, re-identification risk assessments must be undertaken on a case-by-case basis, considering all known the contextual peculiarities of the data in question. Accordingly, the fact that other anonymised datasets had previously been released without issue should have been considered of limited relevance as to whether any issues were likely to arise in relation to the PACE data.

In a similar vein, the FTT's conclusion that QMUL sharing PACE data with a restricted range of independent scientists for research purposes represented an acknowledgement that the data had been anonymised, and therefore was suitable for public disclosure under the FOIA, is also difficult

to understand. Once again, as the UKAN Framework convincingly explains, the motivations and identities of the recipients of anonymised datasets will have a profound impact on the level of risk of re-identification associated with the data contained therein.[88] By the same metric the utilisation of contextual controls, such as confidentiality agreements and data use licences, are also likely to be highly relevant to determining levels of re-identification risk. In other words, just because the disclosure of a dataset to one recipient can be considered low risk, it does not hold to reason that disclosure to another recipient should be considered the same. As has been argued convincingly elsewhere, for instance, an analysis of the GDPR's provisions relating to the definition of personal data and the concept of identifiability reveals that whilst data might validly be thought of as personal in the hands of one party, it does not follow that the same data must necessarily also be considered personal in the hands of another.[89]

For instance, a small office of university researchers seeking to use a dataset for a specific research purpose and operating under a licence to use the dataset only for that specific purpose, might plausibly represent a considerably lower risk than a large multinational corporation, not bound by any licensing agreement, that wished to use the dataset for the purposes of pursuing various unspecified commercial activities for the sake of financial gain. These are issues that have also been noted in both the literature pertaining to the anonymisation of personal data and data protection,[90] as well as risk research and risk management more generally.[91] It is therefore troubling that the FTT appeared to take the attitude that the fact that disclosure to a selected and limited range of independent scientists under controlled conditions was indicative of the PACE data being of a low-risk of re-identification in a broader sense, therefore rendering them suitable for unrestricted public disclosure.

4.4. Looking to the future – the need to devise a new approach to data disclosures

For the reasons outlined above, it can convincingly be argued that the approach of the FTT in the QMUL case was clearly detached from contemporary and authoritative understandings of anonymisation. However, the FTT's approach to data disclosures did broadly coalesce with the FOIA's "release and forget" ethos. As alluded to above, having decided that the PACE data had been anonymised the FTT had no option but to order their disclosure on a "release and forget" basis, such is the wording of the FOIA. The QMUL case not only highlights, therefore, how the courts appear to have a poor grasp of anonymisation, but that they are operating under a legal framework which is also disconnected from anonymisation's practical realities.

As noted previously, the FOIA's approach means that in many circumstances personal data cannot be disclosed following a FOI request, but non-personal data, or personal data that have been successfully anonymised, can be released without restriction and effectively forgotten about. This approach appears to rely upon the assumption that data must always exist in finite, definitive, and unalterable states (i.e. personal data will always be personal, data that are not personal will never be personal and, vice versa, data that have been anonymised will forever remain anonymous). This is evidently not the case. Questions regarding whether certain information constitutes personal data, or whether personal data have been anonymised, are highly contextual, and their answers can, and in many instances will, change drastically with the passage of time. There is no guarantee that data that can be considered anonymous in a certain context one day should necessarily be considered anonymous in that same context the next. Given the possible harms that may be experienced by individuals in the event of their anonymised personal data being "de-anonymised", this presents a severe problem. Whilst it may remain appropriate for some data types to be

disclosed without issue under a "release and forget" approach, this will clearly not be true for all types of data. Accordingly, is simply unrealistic for releasers and recipients of anonymised datasets to continue to remain free from post-release obligations and controls. Accordingly, the FOIA's approach to data disclosures is evidently in need of a fundamental rethink, and a new approach which incorporates such obligations and controls is needed.

The challenge of course, is working out how this might best be done. As has been noted elsewhere, for instance, though debates surrounding issues relating to anonymisation and the regulation of data disclosures, and data-handling practices more generally, have been discussed in the academic and scholarly literature for the best part of a decade, and have even reached the mainstream in some instances, law and policy has not meaningfully moved forward.[92] As a means of attempting to advance the debate in this area, the remainder of this article is dedicated to considering whether the utilisation of data protection by design strategies may help improve upon the abovementioned deficiencies of the FOIA's "release and forget" approach to disclosing data.

5. Data protection by design: a way of reconciling anonymisation and FOIA disclosures?

The term privacy by design entered use around 2000, with the Workshop on Freedom and Privacy by Design at the Computers, Freedom and Privacy 2000 conference,[93] as well as being mentioned in a variety of other academic papers published around the same time.[94] The notion of data protection by design, as an offshoot of privacy by design, is essentially directed at information systems development, with the aim of ensuring that privacy and data protection-related interests are accounted for (i.e. "built in") during the lifecycle of such development.[95] The rationale behind this ethos is the belief that building data protection principles into the architecture of information systems will improve the principles' traction.[96]

Over the last ten or so years talk of privacy and data protection by design has become a staple part of the data protection discourse, and recognition of its significance and potential value has gradually increased. In 2010, for instance, the 32nd International Conference of Data Protection and Privacy Commissioners unanimously passed a resolution recognising privacy by design as an essential component of fundamental privacy protection.[97] Similar sentiments have also more recently been expressed by the Article 29 Working Party,[98] the US Federal Trade Commission, [99] the European Court of Human Rights,[100] and the Court of Justice of the European Union. [101]

As has been noted elsewhere, regulatory approaches to privacy and data protection by design have invited new and innovative new approaches to privacy and data protection rule-making.[102] Perhaps the best examples of this happening in practice are articles 25(1) and 25(2) of the GDPR, which impose upon data controllers an obligation to implement technical and organisational measures designed to implement data-protection principles in an effective manner, integrate safeguards into the processing of personal data which give effect to the other provisions of the GDPR, and to ensure that by default only personal data that are necessary for a specific purpose are processed.[103]

In a sense, it is perhaps strange that the GDPR places such emphasis on the notion of privacy and

data protection by design as these are notions that emphasise the role of the designer and integrator of a system. This is because in many instances the data controllers to whom the GDPR applies will not be responsible for designing their systems as this will be done by a third party. Nevertheless, it is the data controller, rather than the system designer, who will bear the burden of meeting the GDPR's data protection by design obligations. System designers are not formally recognised by the GDPR in any capacity.[104] Perhaps for reasons such as this Art. 25 has been identified as being one of the most unorthodox and ambitious of the GDPR's new provisions.[105] Some, for instance, have expressed doubts about the likelihood of attempts to instil data protection by design ideals via legislative provisions succeeding,[106] whilst others, more generally, have suggested that such initiatives may in fact be counterproductive in some instances, and in others may even create negative privacy impacts.[107] Others, however, have been more optimistic, and have highlighted how privacy by design initiatives have to date enjoyed successes in a number of different areas of application.[108]

Irrespective of its success, or lack thereof, in other areas, however, it would seem data protection by design could potentially play an important role in the context of FOI legislation, specifically in terms of addressing the abovementioned concerns associated with the FOIA's "release and forget" approach to data disclosures. One notable possibility would be the introduction, or "building in", of new organisational measures to the procedure for making FOI requests. A proposed new model built around various organisational measures, including the screening of FOI requests and the utilisation of data licensing mechanisms, is sketched in the following section of the article.

6. A new model for data disclosures made under the FOIA 2000

The authors' proposed new model for data disclosures made under the FOIA represents an alternative to the current "release and forget" approach, is designed to bring the law into alignment with the practical realities of anonymisation, and is comprised of four individual stages:

- A preliminary screening stage;
- An applicant motivation screening stage;
- A risk analysis stage; and
- A licensed disclosure stage.

6.1. Preliminary Screening Stage

In the first stage of the disclosure process an individual making a FOI request would be obligated to state the data/information they were requesting access to, the reasons behind their request, and the purposes for which the requested data would be used in as much detail as possible (e.g. for medical research, monitoring possible corruption, checking government spending etc.). In addition to this, the applicant would be required to explain how, and under what conditions, they would store the data were they to receive them (i.e. whether any security measures would be in place etc.).

The public authority would then be required to assess whether information requested under an FOI request were either about an individual or could theoretically be used to identify an individual and, if so, whether their disclosure would breach the data protection principles as contained within the Data Protection Act 2018.[109]

In other words, the public authority would be required to examine whether the information requested were either personal data or anonymised data that would constitute personal data were they to be de-anonymised, and whether the disclosure of such information would be either be unlawful or unfair.

If the requested information was neither personal data nor anonymised personal data, assuming no other restriction or exemption contained within the FOIA applied, the public authority would be free to disclose the requested information on a "release and forget" basis and would not be required to continue with the subsequent stages of the disclosure process.

If the requested information was personal data, the public authority would be required to assess whether the disclosure of those data would amount to a breach of the data protection principles (i.e. whether the release of those data would be fair and lawful) by way of reference to the relevant provisions of the General Data Protection Regulation, DPA 2018, and the guidance of the ICO.

If the disclosure of the personal data was neither unlawful nor unfair there would be no issue with the release of those data, as such a release would not breach any of the data protection principles. The public authority would once again be free to disclose the requested information on a "release and forget" basis and would not be required to continue with the subsequent stages of the disclosure process. If, however, it was believed that the disclosure of the requested information would be either unlawful or unfair, then the general disclosure of the requested data would not be permissible, and the public authority would be required to anonymise the requested data to the best of their ability and proceed to the next stage in the disclosure process.

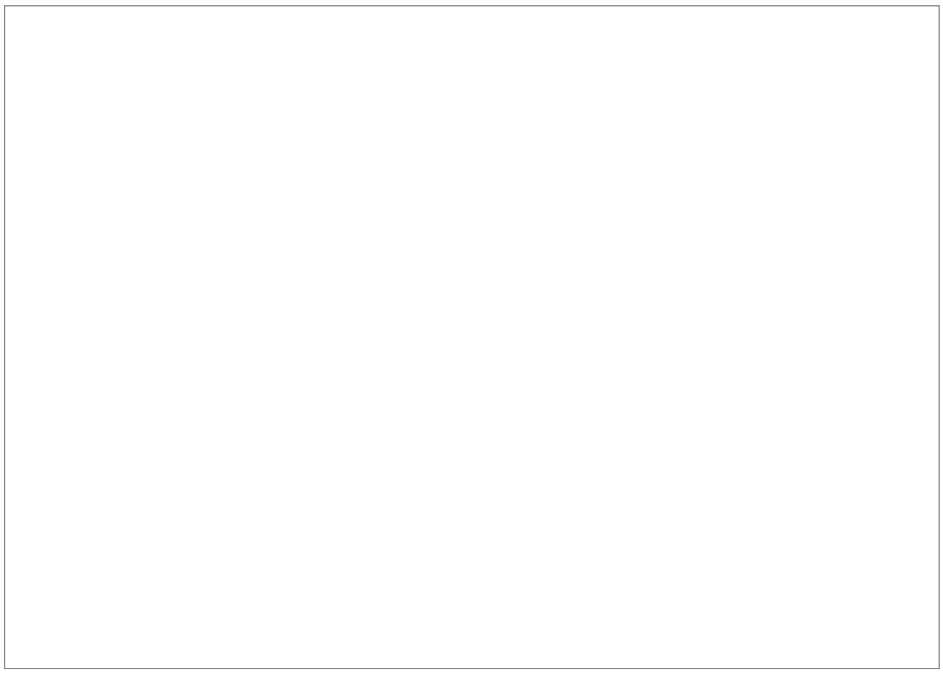
If the requested information was anonymised personal data, the public authority would be required to assess whether the disclosure of those data, were they to be de-anonymised, would amount to a breach of the data protection principles. If the answer to this question was no, then the requested information could be disclosed. If the answer to this question was yes, the public authority would be required to continue with the disclosure process.

Put simply, this stage of proceedings is designed to ensure that only FOI requests that may potentially lead to a breach of an individual's data protection rights are subject to the full disclosure process outlined below. If there is no prospect of an individual's data protection rights being affected, subjecting the requested information to the below risk analyses becomes unnecessary. The inclusion of this stage in the process should, therefore, prevent public authorities from having to undertake superfluous, and possibly cost intensive, risk analyses in relation to all FOI requests.

As noted above, however, as has been argued elsewhere, in most circumstances disclosures of personal data under the FOIA are unlikely to be considered fair or lawful, which would necessitate the utilisation of the subsequent stages of the disclosure process in many cases.[110]

Fig.1, below, shows an explanatory flowchart which illustrates how part one of the preliminary screening stage would operate in practice.

Fig.1: Preliminary Screening Assessment Flowchart



6.2. Applicant Motivation Screening Stage

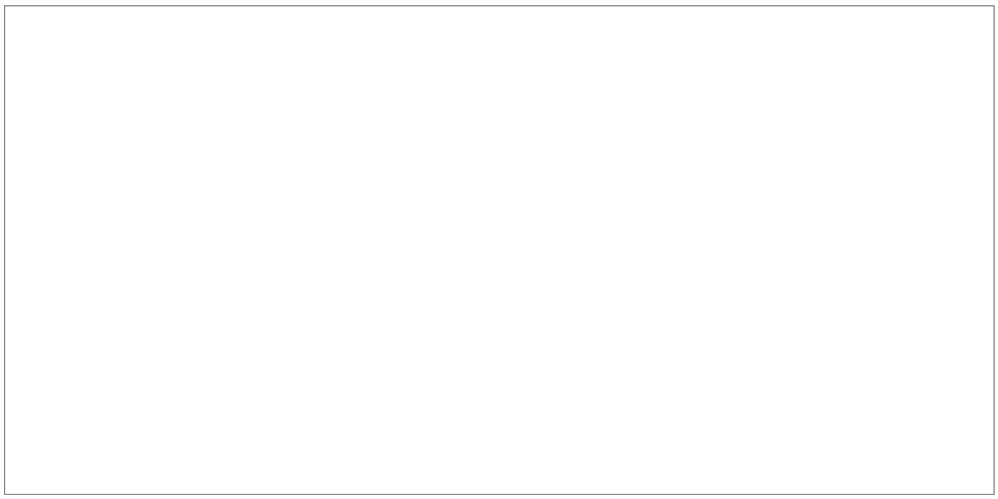
In the second stage of the disclosure process, the public authority would be required to consider the motivations of the individual making the FOI request (which had been provided at the time of the request was made), and the reasons/intended purposes behind their request.

In the perhaps unlikely event an individual was to state a reason that was illegitimate or otherwise incompatible with the thrust of data protection law and freedom of information law (e.g. a desire to obtain personal data, uncover state secrets, or access information received in confidence) their request could be immediately declined. In the more likely event of an individual's request being based on legitimate reasons, the public authority of whom the FOI request was made could proceed to the second stage in the disclosure process.

The inclusion of this applicant screening stage would help public authorities to identify precisely which data were the focus of an individual's FOI request, help with the identification of precisely which data would be necessary to respond to the request in full, help with the making of determinations as part of the risk analysis stage of the disclosure process (outlined below), expedite the rejection of vexatious and/or illegitimate requests, and ensure that the time elapsed between a request being made and the requested data being disclosed was kept to a minimum.

Fig.2, below, shows an explanatory flowchart which illustrates how this stage of the disclosure process would operate in practice.

Fig.2: Applicant Motivation Screening Stage Flowchart



6.3. Risk Analysis Stage

Having identified the data requested by an applicant, the public authority of whom the request had been made would then subject the data requested to up to three different types of risk assessment, the outcomes of which would be measurable on a numerical scale (1 = negligible, 2 = moderate, 3 = high). To be clear, it would be the public authority of whom the FOI request had been made who would bear the burden for carrying out these initial assessments, as is the case at present under the 'release and forget' model. The decisions of public authorities could be appealed to the Information Commissioner, and subsequently to the First Tier Tribunal.

1) The first analysis would require the public authority to assess the risk of the data requested being used to identify an individual were the data to be released openly, without restriction.

If, following this analysis, the public authority considered the risk of identification/re-identification to be negligible (i.e. given a risk score of 1), it could then proceed to the third type of risk analysis listed below.

If, however, the level of risk was deemed to be greater than negligible (i.e. given a score of greater than 1), the second type of risk analysis, also mentioned below, must first be undertaken, and the score from this second risk analysis used in lieu of the risk score obtained from the first analysis.

2) The second analysis would require the public authority to consider the risk of the data requested being used to identify an individual were the data used only and exclusively for the purposes stated in the applicant's initial FOI request, and stored in the way the applicant had stated, as opposed to the data being released completely openly. The more detail the applicant has provided, the easier it would likely be for this analysis to be carried out.

Once a score had been given in relation to this assessment the public authority could then move to the third risk analysis.

There would be no need for a public authority to undertake this second analysis if the complete and unfettered disclosure of the data requested was negligible, as the 'open' release of the data would obviously be broad enough to encompass any specific intended purposes of uses of the applicant.

3) The third and final risk analysis would require the public authority to assess, irrespective of the likelihood of the data being used to identify an individual, the level of severity and/or harm (i.e. impact it could have on an identified individual) were identification to occur.

If difficulties are experienced in calculating the extent of possible harms, the sensitivity of the data should act as a general guide. As has been noted elsewhere, for instance, the more sensitive data are, the greater the harm is likely to be for any affected individuals in the event said data are used for illicit or nefarious purposes.[111]

Having then undertaken the abovementioned exercises in risk-analysis, the public authority would then combine the risk severity score from the first or second risk analyses with the score from the third risk analysis. In so doing, the public authority could discern an overall risk score associated with the disclosure of the data subject to the FOI request. This score could then be used to plot the overall level of risk associated with the disclosure of the data, a concept henceforth referred to as "holistic risk". Once a holistic level of risk had been determined with a prospective data disclosure, the public authority could then move to the third and final stage of disclosure.

6.4. Licensed Disclosure Stage

The final stage in the disclosure process would be built around a system of data licensing. Depending on the level of holistic risk associated with the disclosure of specific data (calculated during the risk analysis stage of the process, outlined above) a public authority would then decide upon an appropriate course of action for those data. Such courses of action might conceivably involve completely open disclosure, restricted disclosure, or even non-disclosure.

In general terms, the greater the level of holistic risk associated with data at the heart of an FOI request, the more likely it would be that the disclosure of those data would be subject to post-release licensing conditions, controls and obligations, which would apply both to the recipient(s) of the data and the releasing public authority.

The recipients of released data, for instance, were such data of a certain risk status, could be placed under licensing obligations to not use data disclosed to them for certain purposes, or to not share those data with any other parties. Concurrently, the releasing public authority could be placed under an obligation to monitor the ongoing situation of those data and, in the event a change in the external environment of those data were to occur and raise the risk status of those data, be obligated to contact the recipient to cease processing of the data, to compel the recipient to destroy all copies of the data, to contact any affected individuals who may be adversely affected, and to act to mitigate any possible emergent harms.

6.5. Discussion

As highlighted previously, the releasing of anonymised personal data under the current "release and forget" model of the FOIA cannot, in most circumstances at least, be considered appropriate, as complete infallible anonymisation is not possible. The unfettered release of such data could, therefore, plausibly give rise to a high level of risk of re-identification and affected individuals suffering consequent harms. However, for the reasons stated previously, as anonymisation is highly contextual it may be possible for such data to be released in such a way that allows for functional anonymisation to be achieved, thereby rendering the limited or controlled disclosure of those data permissible. The use of a data licensing system in conjunction with an obligation for makers of FOI requests to state their reasons and/or motivations could be key to ensuring

functional anonymisation in some contexts. The proposed new model sketched above illustrates how such an approach could work in practice.

For instance, under the proposed model, if data requested via FOI legislation either carried with them a high risk of re-identification were they to be released completely publicly or openly, or the purposes for which the data were requested were likely to give rise to a high-risk of re-identification, then it may be that they could instead be released under a specific licence which would prohibit certain subsequent sharing or usages of those data and mandate certain types of security measures.[112] So long as the terms of such a licence were respected, the data requested could potentially be considered, and remain, functionally anonymised, and so whilst their open release may not be appropriate, their disclosure on a limited licensed basis this may not give rise to a unacceptable level of risk. Obligations could then also be placed on the releaser of data following an FOI request that compelled them to monitor the ongoing data situation of those data in the event of any post-release changes which may impact the susceptibility of those data to re-identification, allowing for the taking of post-release action to mitigate any possible future harms.

Whilst various observers have examined the general possibility of data licensing agreements being used in data protection contexts, it appears little attention has hitherto been specifically devoted to the idea of developing a licensing system for the disclosure of anonymised data pursuant of FOI requests.[113] As shown above, however, it is not especially difficult to envisage how a licensing system for data disclosures built around the notions of anonymisation and risk, where different post-release obligations, would be attached to data depending on their risk of re-identification, might be constructed.

How the proposed system might operate in practice is illustrated in the below hypothetical example:

Person A makes an FOI request to a public authority for dataset X (which is comprised of anonymised personal data) and is required to state the purposes behind their request. Person A states that they wish to acquire the data in dataset X for the purposes of Y.

The public authority of whom the request is made considers that the data contained within dataset X would likely be re-identified if they were to fall into the hands of Person B, or if the data were used for purpose Z.

The release of dataset X under the “release and forget” model would clearly be inappropriate, as releasing the data in this way would mean the data would likely end up in the possession of Person B and/or end up being used for purpose Z.

The public authority, however, considers the probability of the data being re-identified were they to remain in the hands of Person A and used only for purpose Y would likely be negligible with specified security measures. Accordingly, were the data to be disclosed to Person A alone, and used only for purpose Y, dataset X could be considered functionally anonymised.

Pursuant of this, having been made aware of Person A’s intentions, so to ensure that the data remained functionally anonymised, the public authority could release dataset X to Person A under a licence which specified Person A was not permitted to share the data with any other parties (or perhaps specifically not with Person B, or anyone who would likely share the data with Person B), and not to use the data either for any purposes incompatible with purpose Y or for any activity in pursuit of purpose Z.

The public authority itself would then be placed under an obligation to monitor the data situation

of the released data to ensure that in the event any changes might arise that could alter or raise the risk of re-identification associated with those data. If such a situation were to arise, the public authority would be under an obligation to take steps to mitigate any possible harms.

This hypothetical scenario highlights the incorporation of data protection by design initiatives could help reconcile the practical realities of anonymisation with FOI legislation. Notably, the introduction of such a system would go some way to addressing some of the perceived shortcomings of the "release and forget" model of disclosure, specifically the fact that under this existing approach the recipient of data following a successful FOI request is under no specific obligations regarding future sharing and re-uses of the data, meaning that any data disclosed pursuant of an FOI request can effectively become open data, and the fact that the releaser has no obligations post-release to monitor the situation of those data.

Under a model in the vein of that sketched above, however, data could conceivably be released to the makers of FOI requests, who would then be prohibited from any sharing or subsequent uses of the data that could bring the risk of re-identification associated with those data to an unacceptable level. The releasing public authority would also have its own post-release obligations to monitor the ongoing situation of the data and be obligated to act to mitigate any emergent harms. In other words, the use of a model such as that described above would allow for decisions to be made regarding the disclosure of anonymised data on a case-by-case basis, with full consideration being given to the external environment, or "data situation" of the data requested, and how it might change over time post-release. In so doing it would allow the law to treat anonymisation as the ongoing, dynamic and highly-contextual process that it is, rather than a one size fits all, one-time event, and could help bridge the abovementioned disconnect between law and technology.

7. Conclusion

As noted at its outset, the objective of this article was twofold. The first objective was to highlight how anonymisation appears to currently enjoy an uneasy relationship with the law, how the "release and forget" ethos of the FOIA is not fit for purpose, divorced from authoritative contemporary understandings of the concepts of anonymisation and personal data and, more generally, to highlight how a new approach to regulating FOI data disclosures was necessary. The second was to consider whether an alternative approach to data disclosures build around data protection by design elements could potentially help to improve the current situation.

In relation to the first objective, the article explained how answers to questions regarding whether data can be considered either anonymous or personal will depend heavily on context and will be influenced by several factors, such as the nature and character of the data, the existence, identity and desires of any illicit actors or "motivated intruders", and the peculiarities of the "data situation" of those data. Furthermore, it was explained that the data situation of any data will be capable of changing, and likely will change over time. Accordingly, anonymisation must be considered an ongoing and dynamic process that should never necessarily be thought of as being concluded. For these reasons, there are reservations to be had as to both the approach of UK courts to issues relating to anonymisation, and the "release and forget" approach to data disclosures of the FOIA, which cannot be considered fit for purpose in the context of datasets made up of anonymised personal data.

In relation to the article's second objective, the article argued how the incorporation of data by design strategies in the form of an built in obligation for any individual making an FOI request to

declare the reasoning and motives behind their request and a description of their data processing environment, including related security measures, and the linkage of this to a system of data licensing, represents a new approach to regulating FOI requests that could potentially reconcile FOI legislation with the abovementioned issues relating to anonymisation. The alternative model for disclosing data pursuant of an FOI request presented above would allow for datasets consisting of anonymised personal data to be released in a way that prevented, or at least mitigated, possible threats of re-identification and harm that would otherwise be present were the data to be released and forgotten. The result of this would be to allow for data to be shared for the purposes of FOI, whilst simultaneously allowing for the law to engage with the practical realities of anonymisation and afford protection to individuals' whose data are contained within anonymised datasets. In this sense, the model proposed above could have the potential to reconcile FOI legislation and anonymisation.

---

[1] Lecturer in Law, University of Portsmouth, Portsmouth Law School, Richmond Building, Portland Street, Portsmouth, Hampshire PO1 3DE. Email: henry.pearce@port.ac.uk

[2] Professor in Information Technology Law and Data Governance, Southampton Law School, Southampton, Hampshire, SO17 1TR. Email: s.stalla-bourdillon@soton.ac.uk

[3] [hereinafter FOIA]

[4] See, for example: Narayanan, A. and Shmatikov, V. (2008) "Robust De-anonymization of Large Sparse Datasets", *IEEE Symposium on Security and Privacy*; Narayanan, A. and Shmatikov, V. (2009) "De-anonymizing Social Networks", *IEEE Symposium on Security and Privacy*; Ohm, P. (2010) "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization", *UCLA Law Review*; Wondracek, G. et al. (2010) "A Practical Attack to De-anonymise Social Network Users", *IEEE Symposium on Security and Privacy*.

[5] See: Aldhouse, F. (2014) "Anonymisation of Personal Data: A Missed Opportunity for the European Commission", *Computer Law & Security Review*; Stalla-Bourdillon, S. and Knight, A. (2017) "Anonymous Data v. Personal Data – A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data", *Wisconsin International Law Journal*

[6] See, for example: Ohm, P. (n.2)

[7] *Queen Mary University of London v (1) The Information Commissioner and (2) Alem Matthees*, EA/2015/0269

[8] [hereinafter FTT]

[9] [hereinafter UKAN]

[10] Elliot, M. et al. (2016) *The Anonymisation Decision-making Framework*, UKAN.

[11] See, for example: Saunders, B. Kitzinger, J. and Kitzinger, C. (2015) "Anonymising interview data: challenges and compromise in practice", *Qualitative Research* 15(5), pp.616-632. Cavoukian, A. and El Emam, K. (2011) "Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy", available at:

https://fpf.org/wp-content/uploads/2011/07/Dispelling%20the%20Myths%20Surrounding%20De-identification%20Anonymization%20Remains%20a%20Strong%20Tool%20for%20Protecting%20Privacy.pdf

[12] For an overview of the events leading to the enactment leading to the FOIA see: Wood, S. (2003) "From the Hutton Enquiry to driving test routes. The UK Freedom of Information Act (2000): implications for information provision in the UK", *Legal Information Management.*

[13] [hereinafter ECHR]

[14] Art.10 ECHR. As indicated by the European Court of Human Rights, the right to freedom of expression under Art.10 ECHR is also capable of extending to the ability to seek, receive and impart information. See: *Leander v Sweden*

(1987) 9 EHRR 433; *Tarasag v Hungary* (2011) 53 EHRR 3; and *Kenedi v Hungary* (31475/05) (2009) 27 BHRC 355; *Magyar Helsinki Bizottság v Hungary* (18030/11) (2016) ECHR 975.

[15] Liu, M. (2006) "Transparent government and the Freedom of Information Act 2000", *Coventry Law Journal.* It is also worth noting that despite the FOIA enjoying widespread support, in some sectors debates have arisen in respect of its general desirability and effectiveness. See, for instance: Cooper, D. (2005) "UK Freedom of Information Act 2000: boon or bane?", *Company Lawyer*; Worthy, B. (2010) "More Open but Not More Trusted? The Effect of the Freedom of Information Act 2000 on the United Kingdom Central Government", *Governance: An International Journal of Policy, Administration and Institutions* 23(4), pp.561-582; Worthy, B. (2013) ""Some are More Open than Others": Comparing the Impact of the Freedom of Information Act 2000 on Local and Central Government in the UK", *Journal of Comparative Policy Analysis: Research and Practice* 15(5), pp.395-414; Roberts, A. (2002) "Less Government, More Secrecy: Reinvention and the Weakening of Freedom of Information Law", *Public Administrative Review* 60(4), pp.308-320; Hayes, J. (2009) "A Shock To The System: Journalism, Government and the Freedom of Information Act 2000", *Reuters Institute for the Study of Journalism Working Paper*, available at: https://www.ucl.ac.uk/constitution-unit/events/events-archive/Hayes_Working_Paper_Shock_To_The_System.pdf

[16] See, for instance, Section 6 of the Human Rights Act 1998 (HRA), where it is specified that in addition to any court or tribunal "any person certain of whose functions are functions of a public nature" will be considered public authorities. As a general matter, it has been left for the courts to determine the identities of public authorities for the purposes of the HRA. See: *Aston Cantlow Parochial Church Council v Wallbank* [2003] UKHL 37; *YL v Birmingham City Council* [2007] UKHL 27.

[17] See: Section 3 and Schedule 1 FOIA 2000.

[18] Section 10 FOIA.

[19] Other areas of law may indirectly prohibit uses of information disclosed via a FOI request. For example, were the recipient of information following a successful FOI request to use that information to instigate an act of fraud, their actions would likely be unlawful as per the substantive provisions of the Fraud Act 2006.

[20] See: Salganik, M. (2017) *Bit by Bit: Social Research in the Digital Age*, *Princeton,* New Jersey: Princeton University Press, pg.313; Naranayan, A. and Shmatikov, V. (2010) "Myths and fallacies of "Personally Identifiable Information", *Communications of the ACM* 53(6) pp.24-26; Rubinstein, I. and Hartzog, W. (2016) "Anonymization and Risk", *Washington Law Review* 91; Stalla-Bourdillon, S. and Knight, A. (n.3); Ohm, P (n.2)

[21] At this stage the authors would like to note that some public authorities voluntarily publish all the data they disclose pursuant of FOI requests on disclosure logs on their website, which might be described as attempts to "remember". This, the authors contend, is commendable and is a practice that should be encouraged. In this regard it is also worth noting the Re-use of Public Sector Information Regulations 2015 also require public sector bodies to produce information asset lists, containing the main information (including that which is unpublished) that they hold in relation to the performance of their public tasks. This again, could be used to illustrate the existence of some public-sector initiatives geared towards remembering. It is, however, important to emphasise the fact that the FOIA 2005 itself does not impose any post-release obligations on public authorities which require them to behave in such a way, hence why it, as a legislative instrument, can be said to endorse a "release and forget" ethos. It is also worth noting that on 25th April 2018 the European Commission adopted a proposal for a revision of the PSI Directive, which the Re-use of Public Sector Information Regulations 2015 transposes into English law, that aims to facilitate the creation of a common data space in the EU. See: European Commission (last accessed December 2018) "Proposal for a revision of the Public Sector Information (PSI) Directive", available at:

https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive. On the re-use of public sector information and its intersection with data protection law generally, see: Article 29 Working Party (2013) *Opinion 06/2013 on open data and public sector information ('PSI') reuse,* 1021/00/EN WP207, available at: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf.

[22] See: ss.23-27 FOIA. Interestingly, and as has been noted elsewhere, the exemptions listed under Part II FOIA are considerably more extensive and wide-ranging than many any other freedom of information regimes found in other common law jurisdictions. See also: Liu, M. (n.13).

[23] See Sections 40 (1)-(3) FOIA 2000.

[24] [hereinafter DPA 1998]

[25] Section 1(1) Data Protection Act 1998.

[26] [hereinafter DPA 2018]

[27] Regulation (EU) 2016/679 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 [hereinafter GDPR] At this point it should be noted that despite "Brexit", the UK government has signalled its intentions for the substance of the GDPR, if not the GDPR itself, will be a part of UK law after the UK leaves the European Union in March 2019. However, due to the GDPR coming into force on 25th May 2018, its status as an EU Regulation means it will be directly applicable in the UK until this occurs.

[28] Sections 3 (2)-(3) Data Protection Act 2018.

[29] Section 86 Data Protection Act 2018.

[30] Section 87 Data Protection Act 2018.

[31] Section 88 Data Protection Act 2018.

[32] Section 89 Data Protection Act 2018.

[33] Section 90 Data Protection Act 2018.

[34] Section 91 Data Protection Act 2018.

[35] For a detailed explanation of the different legitimising grounds for the processing of personal data under the GDPR see: Stalla-Bourdillon, S. Pearce, H. and Tsakalakis, N. (2018) "The GDPR, a game changer for electronic identification schemes? The case study of Gov.UK Verify", *Computer Law & Security Review* 34(4), pp.784-805.

[36] McCluskey, C. (2017) "How will the GDPR affect FOI law?", *Freedom of Information* 13(5)

[37] See: GDPR Art.6(1)(f), which states: "*Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks*." An explanation for this restriction is provided in Recital 47, which reads: "*Given that it is for the legislator to provide by law the legal basis for public authorities to process personal data, that legal basis should not apply to the processing by public authorities in the performance of their tasks.*"

[38] See paragraphs 58 and 307 of Schedule 19 DPA 2018.

[39] [hereinafter ICO]

[40] ICO (2018) *Requests for personal data about public authority employees: Freedom of Information Act, Environmental Information Regulations*. Available at: https://ico.org.uk/media/for-organisations/documents/1187/section_40_requests_for_personal_data_about_employees.pdf

[41] This data, according to the ICO, is likely to relate to the most personal aspects of individuals' lives, such as their health and sexual life. *Ibid.*, pg.6. See also, Schedule 10 DPA 2018, which states that data pertaining to an individual's racial and ethnic origins would also likely constitute sensitive personal data.

[42] ICO (n.39) pp.7-10.

[43] See, for instance: *Edem v IC & Financial Services Authority* [2014] EWCA Civ 92. For other recent examples of cases involving similar issues see: *Grant Workman v Information Commissioner & Home Office* [2017] UKFTT EA/2017/0127; *Carole Evans v Information Commissioner and another* [2016] UKFTT EA/2016/0131; *Janet Dedman v Information Commissioner* [2017] UKFTT EA/2016/0142; and *Thompson v Information Commissioner and another* [2016] UKFTT EA/2016/0044.

[44] Recital 26 GDPR.

[45] Though in practical terms anonymisation spans a great variety of techniques, it can generally be described as a process through which information within a dataset can be manipulated to make it more difficult, or impossible, to identify individuals from that information. Generally speaking, these techniques can be split into two categories: "randomisation" techniques which alter the veracity of data, such as noise addition, permutation and differential

privacy, and "generalisation" techniques which reduce the granularity of data, such as K-anonymity and L-diversity. For an overview of a range of notable anonymisation techniques, see: Article 29 Working Party (2014) *Opinion 05/2014 on Anonymisation Techniques,* 0829/14/EN WP216, available at: http://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf. See also: Sweeney, L. (2002) "K-Anonymity: A Model For Protecting Privacy", *International Journal Of Uncertainty, Fuzziness and Knowledge-Based Systems*; Machanavajjhala, A. et al. (2007) "L-Diversity", *ACM Transactions On Knowledge Discovery From Data*; Saxby, S. Knight, A. and Pearce, H. "Piercing the Anonymity Veil: Re-identification Risk and the UK Transparency Agenda" in *Information Ethics and Security: Future of International World Time*, ed. by Kierkegaard, S. (2014) International Association of IT Lawyers, pg.6.

[46] On big data, see: Mayer-Schonberger, V. and Cukier, K. (2013*) Big Data: A Revolution That Will Transform How We Live Work and Think*, Oxford: John Murray Publishers; Fishleigh, J. (2014) "A non-technical journey into the world of Big Data: an introduction", *Legal Information Management*; boyd, d. and Crawford, K. (2012) "Critical Questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon", *Information, Communication and Society*

[47] Seminally, in 2008, Arvind Narayanan and Vitaly Shmatikov of the University of Texas demonstrated that by applying a bespoke de-anonymisation methodology to the Netflix Prize dataset,[47] which contained the supposedly anonymised film ratings of 500,000 Netflix subscribers – the world's largest online film rental platform – it would be possible for an adversary who possessed minimal knowledge about an individual Netflix subscriber to easily identify the record of that individual in the database. By using the Internet Movie Database as a source of background information, Narayanan and Shmatikov were able to successfully identify several Netflix records of known users and, as a result, uncovered those users' political preferences and a range of other sensitive character and identity traits. See: Narayanan, A. and Shmatikov, V. (2007) "How To Break Anonymity of the Netflix Prize Dataset", *The University of Texas At* Austin. See also: Narayanan, A. and Felton, E. (2014) "No silver bullet: de-identification still doesn't work", *Princeton University*. Available at: http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf. Another notable incident was highlighted in a 2013 article published by Harvard University academics. Here, the results of a study were reported, in which it was shown that it had been possible to correctly identify 84-97% of the individuals who had participated in the US' Personal Genome Project by way of fusing supposedly anonymised public profiles (containing zip code, birth date and gender data) with public voter polls and mining for names hidden in attached documents. See: Sweeney, L. et al. (2014) "Identifying Participants in the Personal Genome Project by Name", *Harvard University, Data Privacy Lab. White Paper 1021-1*; The challenges posed to the law by the limits of anonymisation techniques was brought to the attention of legal scholars and practitioners in 2010 by Paul Ohm. Ohm, P. (n.2)

[48] *Ibid*., See also: Aldhouse, F. (n.3); Saxby, S. Knight, A. and Pearce, H. (n.44); Stalla-Bourdillon, S. and Knight, A. (n.3)

[49] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques (European Commission Working Paper No.216, 0829/14/EN, 2014)

[50] *Ibid.*, pp.11-12. For further discussion of these re-identification risks, see: Hu, R. et al. "Bringing Policy, Regulation, and Practice? A Techno-Legal Analysis of Three Types of Data in the GDPR" in *Data Protection and Privacy: The Age of Intelligent Machines"* ed. by Leenes, R. et al. (2017) Oxford: Hart

[51] *Ibid.*, See also: Ohm, P. (n.2) pg.1707.

[52] *Ibid*

[53] See, for example: Cavoukian, A. and El Emam, K. (n.9): Cavoukian, A. and Castro, D. (2014) "Big data and innovation, setting the record straight: de-identification does work", *Office of the Information and Privacy Commissioner, Ontario*. Available at: http://www2.itif.org/2014-big-data-deidentification.pdf; El Emam, K. (2015) "Anonymising and sharing individual patient data", *BMJ*. For an overview of this debate, see: Nunan, D. and Di Domenico, M. (2016) "Exploring Reidentification Risk: Is Anonymisation a Promise we can Keep?", *International Journal of Market Research*, 58(1), pp.19-34.

[54] Recital 26 of the GDPR reiterates that only data that concern an identified or identifiable natural person should be considered "personal" and specifies that "*To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used…to identify the natural person directly or indirectly.*" The clear

implication being that if an individual cannot be identified from data via "means reasonably likely used" the data should not be considered personal.

[55] See, for example: Schwartz, P and Solove, D. (2012) "The PII Problem: Privacy and a New Concept of Personally Identifiable Information", *New York University Law Review*; Kuan Hon, W. Millard, C. and Walden, I. "What is Regulated as Personal Data in Clouds?" in *Cloud Computing Law* (2013) ed. by Millard, C. Oxford: Oxford University Press; Aldhouse, F. (2014) "Anonymisation of Personal Data: A Missed Opportunity for the European Commission", *Computer Law & Security Review*; Stalla-Bourdillon, S. and Knight, A. (2017) "Anonymous Data v. Personal Data – A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data", *Wisconsin International Law Journal*; Esayas, S. (2015) "The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the "all or nothing" approach", *European Journal of Law and Technology* 6(2).

[56] Hu, R. et al. (n.49)

[57] ICO (2012) *Anonymisation: managing data protection risk code of practice.* [hereinafter the code] Available at: https://ico.org.uk/media/1061/anonymisation-code.pdf

[58] *Ibid*

[59] The UK Anonymisation Network (UKAN), established in 2012, and co-ordinated by the University of Manchester, the University of Southampton, the Open Data Institute (ODI) and Office for National Statistics (ONS), is an organisation geared towards establishing best practice in anonymisation, and offers advice and information to those seeking to share personal data.

[60] Elliot, M. et al. (n.8) [hereinafter the UKAN Framework]

[61] On its rear cover, for instance, the Framework's contents are endorsed by a range of significant figures, notably Elizabeth Denham, the UK Information Commissioner.

[62] Elliot, M. et al. (n.8)

[63] *Ibid.*, pg.16.

[64] *Ibid.*, pg.17.

[65] *Ibid.*, pg.15. This is an approach that has been acknowledged and/or endorsed elsewhere. See: Elliot, M. et al. (2018)"Functional anonymisation: Personal data and the data environment", *Computer Law & Security Review* 34(2), pg.204-221; O'Keefe, C. et al. (2017) *The De-identification Decision-Making Framework*, CSIRO Reports EP173122 and EP175702, available at: https://publications.csiro.au/rpr/download?pid=csiro:EP175702&dsid=DS1; Simperl, E. O'Hara, K. and Gomer, R. (2016) *Analytical Report 3:Open Data and Privacy*, European Data Portal, available at: https://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf; March, S. et al. (2015) *Data protection aspects concerning the use of social or routine data*, Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research, available at: http://doku.iab.de/fdz/reporte/2015/MR_12-15_EN.pdf; Stalla-Bourdillon, S. and Knight, A. (n.3) Cavoukian, A. and El Emam, K. (n.9); Saunders, B. Kitzinger, J. and Kitzinger, C (n.9); Roberts, D. (2018) "Why Contextual Privacy Controls Are Essential", *Privitar* (last accessed: 13th November 2018), available at: https://www.privitar.com/listing/contextual-privacy.

[66] *Ibid.*, pg.114.

[67] The idea of releasers of anonymised datasets having obligations to monitor and mitigate risks post-release is an idea that has also gained some support in the academic literature. See, for example: Rubinstein, I. and Hartzog, W. (2016) (n.18); Stalla-Bourdillon, S. and Knight, A. (n.3)

[68] The key tenets of the functional anonymisation approach of the UKAN Framework were reiterated in a recently published paper written by the same authors. See: Elliot, M. et al. (2018) "Functional Anonymisation: Personal data and the data environment", *Computer Law & Security Review* 34, pp.204-221.

[69] Stalla-Bourdillon, S. and Knight, A. (n.3)

[70] [hereinafter CJEU]

[71] See: *Patrick Breyer v Budesrepublik Deutschland, Case C-582/14 [2016] (EU:C:2016:930)*

[72] [hereinafter QMUL]

[73] The acronym, PACE, was derived from the study's formal title, *Pacing, graded Activity and Cognitive behaviour therapy: a randomised Evaluation.*

[74] ICO (2015) Freedom of Information Act 2000 (FOIA) Decision Notice, 27 October 2015. Available at: https://ico.org.uk/media/action-weve-taken/decision-notices/2015/1560081/fs_50565190.pdf

[75] For an overview of these criticisms, see: Geraghty, K. (2017) "Further commentary on the PACE trial: Biased methods and unreliable outcomes", *Journal of Health Psychology* 22(9).

[76] [hereinafter IC] (n.73)

[77] *Ibid*

[78] *Ibid*

[79] *Ibid*

[80] ICO (2012) *Anonymisation: managing data protection risk code of practice*, pg.22.

[81] *Queen Mary University of London v (1) The Information Commissioner and (2) Alem Matthees*, EA/2015/0269

[82] *Ibid*.

[83] As an interesting point of reference, the FTT's decision in the immediate case appeared to echo the decision made in an earlier case in which it was held that recipients of anonymised abortion statistics could not be used to identify individuals, and thus said statistics were not considered personal data under the DPA 1998. *R (Department of Health) v Information Commissioner [2011] EWHC 1430 (Admin)*

[84] Elliot, M. et al. (n.8)

[85] ICO (n.73)

[86] *Queen Mary University of London v (1) The Information Commissioner and (2) Alem Matthees*, EA/2015/0269

[87] This is a position that has recently been endorsed by the CJEU, notably in the abovementioned *Breyer* case. The case also confirmed that in order for an individual to be considered identifiable from data it is not enough for there to be a theoretical chance of their identification, but that their identification, by way of reference to the contextual peculiarities of the data, must be reasonably likely. See above at n.68.

[88] Elliot, M. et al. (n.8)

[89] Mourby, M. et al. (2018) "Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK", *Computer Law & Security Review* 34(2), pp.222-233.

[90] *Ibid*., See also: Pearce, H. (2017) "Big data and the reform of the European data protection framework: an overview of concerns associated with proposals for risk management-based approaches to the concept of personal data", *Information and Communications Technology Law* 26(3), pp.312-335.

[91] See, for example: Black, J. and Baldwin, R. (2010) "Really Responsive Risk-Based Regulation", *Law and Policy* 32(2); Baldwin, R. and Cave, M. (1998) *Understanding Regulation: Theory, Strategy and Practice*, Oxford: Oxford University Press.

[92] Rubinstein, I. and Hartzog, W. (n.18). See also: Craddock, E. Stalla-Bourdillon, S. and Millard, D. (2017) "Nobody puts data in a corner? Why a new approach to categorising personal data is required for the obligation to inform", *Computer Law & Security Review* 33(2), pp.142-158.

[93] CFP2000, Conference on Computers, Freedom & Privacy (2000)

[94] See: Veale, M. Binns, R. and Ausloos, J. (2018) "When Data Protection by Design and Data Subject Rights Clash", *International Data Privacy Law*.

[95] As has been noted in the literature, there are many different strategies and tools that data controllers might deploy in pursuit of this objective. See: Colesky, M. Hoepman, J. and Hillen, C. (2016) "A Critical Analysis of Privacy Design Strategies", *IEE Security and Privacy Workshops*, pp.33-40; van Rest, J. et al. "Designing Privacy-by-Design" in

Privacy Technologies and Policy. APF 2012. Lecture Notes in Computer Science, Vol.8319 (2014) ed. by Preneel, B. and Ikonomou, D. Springer, Berlin; Le Métayer, D. "Privacy by Design: A Matter of Choice" in *Data Protection in a Profiled World* (2010) ed. by Gutwirth, S. Poullet, Y. and De Hert, P. Springer: Dordrecht. pp.323-334; Hoepman, J. "Privacy by Design Strategies" in *ICT Systems Security and Privacy Protection. SEC 2014. IFIP Advances in Information and Communication Technology*, ed. by Cuppens-Boulahia, N. et al. (2014) Springer: Berlin. pp.446-459. See also: Mulligan, D. and King, J. (2012) "Bridging the Gap between Privacy and Design", *Journal of Constitutional Law* 14(4), pp.989-1034.

[96] Cavoukian, A. (2011) "Privacy by Design in Law, Policy and Practice: A White Paper for Regulators, Decision-makers and Policy-makers", *Information and Privacy Commissioner, Ontario, Canada*; Bygrave, L. (2017) "Data Protection by design and by default: Deciphering the EU's legislative requirements", *Oslo Law Review* 4(2), pp.105-120; Hildebrandt, M. and Tielemans, L. (2013) "Data protection by design and technology neutral law", *Computer Law & Security Review* 29(5), pp.509-521; van Rest, J. et al. *Ibid*.

[97] See: Bygrave, L. *Ibid*.

[98] *Ibid*

[99] *Ibid*

[100] The ECtHR has historically made several judgments which have embraced the ideals of privacy/data protection by design. See, for instance, the decision of the Court in *I v Finland. Appl. No.20511/03, Judgment of 17 July 2008*, in which it was held that Finland was in breach of Art.8 ECHR due to a failure to implement technical/operational measures as a means of ensuring the confidentiality of patient medical data in hospitals.

[101] Though the CJEU has not ruled directly on the matter of privacy/data protection by design, in the *Google Spain* case, by rejecting Google's argument that its search engine operations were value neutral robotic applications of algorithms outside the scope of data protection law, the CJEU compelled Google (and other search engine providers) to reconfigure systemic aspects of those operations so that they would be more privacy/data protection friendly. *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González. Case C-131/12 [2014] (EU:C:2014:317)* On this issue, see: Bygrave, L. (n.95).

[102] Cavoukian, A. (2011) "Privacy by Design in Law, Policy and Practice: A White Paper for Regulators, Decision-makers and Policy-makers", *Information and Privacy Commissioner, Ontario, Canada*. pg.17. On the issue of regulating privacy by design more generally, see: Rubinestein, I. (2011) "Regulating Privacy by Design", *Berkley Technology Law Journal*, pp.1049-1456.

[103] As an interesting point of reference, Article 17 of the now defunct Data Protection Directive also appeared to imply a privacy by design requirement which required data controllers to implement appropriate technical and organisational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorised disclosure or access.

[104] van Rest, J. et al. (n.94). pg.57.

[105] Bygrave, L. (n.95)

[106] Koops, B. and Leenes, R. (2014) "Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data protection law", *International Review of Law, Computers and Technology* 28(3), pp.159-171; Spiekermann, S. (2012) "The challenges of privacy by design", *Communications of the ACM* 55(7), ppp.38-40; Shapiro, S. (2010) "Privacy by design: moving from art to practice", *Communications of the ACM* 53(10), pp.27-29.

[107] It has been argued, for instance, that certain confidentiality-focused data protection by design strategies used by large data controllers leave data re-identifiable by capable adversaries while heavily limited controllers' ability to provide data subject rights, such as access, erasure and objection. See: Veale, M. Binns, R. and Ausloos, J. (n.93). See also: Brown, I. (2014) "Britain's Smart meter programme: A case study in privacy by design", *International Review of Law Computers & Technology* 28(2), pp.172-184.

[108] Electronic health cards, electronic ID cards, and electronic proof of earnings, for instance, have been identified as areas of application in which privacy by design has been highly beneficial. See: Schaar, P. (2010) "Privacy by Design", *Identity in the Information Society* 3(2), pp.267-274. Remote healthcare technologies and big data analytics are other areas of application in which the potential for privacy by design has been mooted. See: Cavoukian, A. et al.

(2010) "Remote home health care technologies: how to ensure privacy? Build it in: Privacy by Design", *Identity in the Information Society* 3(2), pp.363-378 and Cavoukian, A. and Jonas, J. (2012) "Privacy by Design in the Age of Big Data", available at: https://jeffjonas.typepad.com/Privacy-by-Design-in-the-Era-of-Big-Data.pdf

[109] See: Sections 86-91 Data Protection Act 2018 [hereinafter DPA 2018]

[110] McCluskey, C. (n.34)

[111] Elliot, M. et al. (n.8). See also: Milne, G. et al. (2017) "Information Sensitivity Typology: Mapping the Degree and Type of Risk Consumers Perceive in Personal Data Sharing", *The Journal of Consumer Affairs* 51(1), pp.133-161.

[112] To this end, such a model would bear some similarities to the Re-use of Public Sector Information Regulations 2005, which permit public sector bodies to impose licensing restrictions on the re-use of information in some situations. See also: Article 29 Working Party (n.19)

[113] See, for example: Verheul, E. et al. (2016) "Polymorphic Encryption and Pseudonymisation for Personal Healthcare: A Whitepaper", *Institute for Computing and Information Sciences, Radboud University Nijmegen, The Netherlands*; Popescu, A. et al. "Increasing Transparency and Privacy for Online Social Network Users – USEMP Value Model, Scoring Framework and Legal" in Privacy *Technologies and Privacy. APF 2015. Lecture Notes in Computer Science, vol 9484*, ed. by Berendt, B. et al. (2015) Cham: Springer, pp.38-59; Joung, Y. and Cha, S. "Online Personal Data Licensing: Regulating Abuse of Personal Data in Cyberspace", in *Intellectual Property Protection for Multimedia Information Technology*, ed. by Sasaki, H. (2007) Hershey, Pennsylvania: IGI Global, pp.165-185; Joung, Y. et al. (2005) "On personal data license design and negotiation", *Computer Software and Applications Conference, 2005. COMPSAC 2005. 29th Annual International*.