Updated to take account of the version of the Bill published March 2023

DATA PROTECTION AND DIGITAL INFORMATION BILL

New definition of personal data

Summary: The Data Protection and Digital Information Bill (the 'Bill') creates a new definition of personal data under the UK GDPR. The new definition is complex, and it is difficult to assess its long-term impact. At a minimum, it will make it somewhat easier for controllers to treat datasets as effectively anonymised, and therefore outside the GDPR's scope. But it might also have unpredictable consequences, increase instances of reidentification from anonymous data by hostile actors, and/or lead controllers to rely on spurious legal arguments in an attempt to take important and consequential types of processing outside the scope of the data protection regime.

The new definition

1. s.1 of the Bill amends s.3 of the Data Protection Act 2018 (the 'DPA') and creates a new s.3A DPA that provide for the definition. These sections set out two cases in which a person is deemed to be identifiable from data (making it personal data):

> *"(2) The first case is where the living individual is identifiable (as described in section 3(3)) by the controller or processor by reasonable means* ***at the time of the processing****.*
>
> *(3) The second case is* ***where the controller or processor knows, or ought reasonably to know****, that*
>
> > *(a)* ***another person will, or is likely to****, obtain the information as a result of the processing, and*
> >
> > *(b) the living individual will be, or is likely to be, identifiable (as described in section 3(3)) by that person* ***by reasonable means at the time of the processing.***

> *(4) For the purposes of this section, an individual is identifiable by a person "by reasonable means" if the individual is identifiable by the person by any means that the person is reasonably likely to use.*
>
> *(5) For the purposes of subsection (4), whether a person is reasonably likely to use a means of identifying an **individual is to be determined taking into account**, among other things—*
>
> > *(a) **the time, effort and costs involved** in identifying the individual by that means, and*
> >
> > *(b) the technology and other resources **available to the person**."* (emphasis added)

2. Under the current regime, the test for whether a person is identifiable from data is most comprehensively stated in Case C-582/14 - *Breyer*, which requires a consideration of "all the means likely reasonably to be used either by the controller **or by any other person**". (emphasis added)

3. Thus, determining identifiability under the revised definition in the Bill is more subjective and fact specific. The Bill proposes that it is not enough that a theoretical individual using the most state-of-the-art reidentification technology could identify someone from data; a specific other person must be had in mind, who is likely to identify an individual, taking into account the resources and technology *actually* available to them.

4. The Bill continues to use the word 'identify', which has been held – most notably in the case of *Vidal-Hall v Google*[1] - to include where an individual is 'singled out':

> *"identification for the purposes of data protection is about data that 'individuates' the individual, in the sense that they are singled out and distinguished from all others. It is immaterial that the BGI does not name the user. The BGI singles them out and therefore directly identifies them."* (at § 115)

[1] [2015] EWCA Civ 311

5. There does not appear to be any intention to change this[2]. Thus the change in definition of personal data should not take processing where individuals are not directly named, but are still singled out – e.g. processing using IP addresses or pseudonymous cookie identifiers – outside the scope of the GDPR. The principal intended effect seems to be to lower the bar for when data can be said to have been truly 'anonymised', and therefore taken outside the scope of the UK GDPR. In turn, there will be more datasets which controllers are able to process (or allow others to process) outside of the scope of the GDPR, provided such processing does not result in any individual being 'singled out', because the bar for those datasets to be 'anonymous' will be lower.

<u>'At the time of the processing'</u>

6. The intention behind the inclusion of the words 'at the time of the processing' in §3A(2) and (3)(b) is not completely clear. The broad definition of processing (Article 4 GDPR) means that most acts in relation to data will be processing. The requirement to consider the potential for identification 'at the time of the processing' is therefore effectively a continuing one for as long as a controller holds the data in question. The new wording does not therefore seem to be intended to limit the controller's obligation to consider identifiability to one point in time – e.g., the point of collection of the data. That position is borne out by the Government's response to the consultation.

7. The Government's consultation response states:

> *"[…] this could be where a living individual is identifiable by the controller or processor by "reasonable means", taking into account, among other things, the* ***technology available at the time of the processing****'* (emphasis added)

8. This explanation suggests that the wording has been added to limit controllers' consideration of identifiability to the means available at the time of the processing. That is, controllers do not need to consider theoretical future means by which they or another person might one day be able to use to identify a living individual from seemingly anonymous data. <u>If that is the intention, the drafting</u>

[2] The relevant section of the government's consultation response is entitled 'clarifying data regarded as anonymous'.

could be improved by substituting the phrase ‘by reasonable means *available* at the time of the processing’ in new §3A(2) and (3)(b) of the DPA. If however there is a different intention, such as limiting the consideration to the point in time of collection of the data, that should be made clear.

Reidentification by hostile actors

9. The change could lead to more instances in which individuals are identified from ‘anonymous’ datasets by hostile actors. Consider a dataset with a small number of data points for each record, which (i) by itself does not enable a specific data subject to be singled out, but (ii) when combined with information from another source (e.g. by a hacker), would enable identification.

10. Under the current regime and the test in *Breyer*, this dataset would likely be treated as personal data: a hacker – even a theoretical one – meets the definition of ‘any other person.’ The controller would therefore be required to protect the dataset under the GDPR.

11. Under the new regime, the controller should first consider (new s.3A(3)(a) of the DPA) whether the data set is ‘likely’ to be obtained by a hacker ‘as a result of [the controller’s] processing’. The risk of a hacker obtaining the dataset may be low; data breaches happen only occasionally. It would be reasonable to say that the information is not ‘likely’ to be obtained by another person (the hacker) as a result of the controller’s processing. The most recent version of new Article 3A provides:

> *“The reference in subsection (3)(a) to obtaining the information as a result of the processing includes obtaining the information as a result of the controller or processor carrying out the processing without implementing appropriate technical and organisational measures to mitigate the risk of the information being obtained by persons with whom the controller or processor does not intend to share the information.”*

12. Whilst this is welcome, it does not substantially alter the analysis, since the key test remains whether the other person is ‘likely’ to obtain information enabling

identification. If a malicious actor is not likely to do so – it is merely possible, or a risk to guard against – then the main dataset in question is not personal data.

13. The new definition of personal data therefore seems to leave such a dataset outside the scope of the GDPR's protection. But the risk of a hacker obtaining these datasets is very real, despite being low. This creates a paradoxical situation in which the controller is not obliged to protect the data – as it is not 'personal data' – despite the very real (if low) risk of a data breach that could lead to reidentification of individuals from the dataset.

14. This analysis could apply to many real-world processing situations. This – presumably unintended – consequence could have a serious impact on individuals' data rights and lead to damaging privacy breaches.

15. This problem could be addressed by amending s.3A(3)(a) to including where there is a material risk of the information being obtained by another person.

Spurious arguments by data controllers

16. Controllers may attempt to argue that processing that is very abstracted from data subject's 'real' or 'civil' identities – e.g., processing for the purposes of behavioural advertising technology ('*ad-tech*') which takes place automatically and uses only pseudonymous identifiers – is no longer in scope of the GDPR as a result of the new definition of personal data. Indeed, controllers have made similar arguments even under the current definition of personal data.

17. Due to the continued use of the word 'identify' and the fact that this includes 'singling out', such arguments appear to have little prospect of success. Again taking the example of *ad-tech*, processing whose very purpose is to take an action in relation to an individual (e.g. by changing their browsing experience) deliberately singles them out and is therefore processing of personal data, including under the new definition.

18. Whilst these arguments may have limited prospect of success before a court, they will need to be strongly resisted – including where they are used outside court (e.g., in refusing to recognise data subject rights).

19. The new definition of personal data may also interact unpredictably with Article 11 GDPR, which permits controllers to refuse the exercise of the data subject rights contained in Articles 15 to 20 where they cannot identify a data subject (which would seem to be easier to demonstrate based on s.1 of the Bill). This could create a barrier to the exercise of data subject rights.

The need to monitor

20. The Government's intentions with this change appear modest. But amending the definition of so fundamental a concept in the data protection regime carries a significant risk of unintended consequences. Advocates for data rights will need to closely monitor:

   i. Whether there is an increase in purportedly anonymised datasets which go on to undergo reidentification under this looser regime; and

   ii. The extent to which controllers begin to place excessive reliance on the new definition, carrying out more hidden processing without safeguards such as transparency in the (misconceived) belief that their processing no longer falls within the scope of the GDPR.