Personal data, anonymisation and pseudonymisation under the GDPR

July 2016

In the 20 years that have passed since the enactment of the Data Protection Directive (the 'Directive'), the volume of, and ease of access to, information about us has increased exponentially. In recognition of this widely accepted but more intrusive fact of virtual life, one of the EU Commission's stated aims in drafting the General Data Protection Regulation (the 'GDPR') was to update and modernise the EU data protection regime to account for new kinds of potentially identifying information. In today's digital world, the GDPR, including questions about the nature of personal data and whether it can anonymised, will therefore continue to be relevant for many organisations, despite Brexit[1].

What changes does the GDPR introduce?

Concept	The Directive	The GDPR (changes in bold)
Personal data	Any information relating to an identified or identifiable natural person.	No change.
Identifiable person	One who can be identified directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.	One who can be identified directly or indirectly, in particular by reference to **an identifier such as a name**, an identification number, **location data, online identifier** or to one or more factors specific to the physical, physiological, **genetic**, mental, economic, cultural or social identity of that person.
Identifiability (explained in Recital 26)	To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person.	To determine whether a person is identifiable, account should be taken of all the means reasonably likely to be used, **such as singling out,** either by the controller or by any other person to identify the individual **directly or indirectly**.

[1] See Slaughter and May's briefing: Brexit and Data Protection: business as usual

This Briefing will look at some of the 'new' categories of personal data the GDPR introduces and will assess the impact of these additions. It will also consider some of the solutions commonly used to remove or reduce the compliance burden on organisations such as anonymisation and pseudonymisation.

How do you identify someone (directly or indirectly)?

Names

The GDPR (which will apply in all Member States from 25 May 2018) confirms that a name can identify an individual, echoing the Court of Appeal's finding in Edem[2] that "A name is personal data unless it is so common that without further information, such as its use in a work context, a person would remain unidentifiable despite its disclosure". Direct identification would usually include a name, with or without contextual information depending on how common the name is. Indirect identification would usually result from a combination of data or identifiers (e.g. the UK Prime Minister). As the ICO has stated: "Simply because you do not know the name of an individual does not mean that you cannot identify that individual" [3].

Location data

Location data is not defined in the GDPR. At its most general, it is likely to be any form of data that has a geographic position associated to it. Examples include information collected by wireless networks, swipe cards and smart mobile devices for a variety of purposes such as navigation, augmented reality, whereabouts tracking and location based advertising, to name but a few.

> The Interactive Advertising Bureau recently stated that "Location data not only improves publishers' own advertising insights but also opens up completely new avenues of monetization", which explains why "location-targeted mobile ad spending is expected to grow from $8.4 billion in 2015 to $11.3 billion in 2016".

The value (and privacy risk) in location data lies in its capacity to infer even more personal data than the face value of the original information – anything from sleeping place to travel patterns and even religion or health issues.

In light of the significant growth of the Internet of Things (IoT) market, MAC addresses (by which a device is commonly identified on a network) are another example that raises concerns from a privacy perspective. They are likely to result in the identification of an individual under the GDPR, especially where they relate to a personal device such as a fitness tracker or a smartphone. Given the 'mission creep' potential of location data and the views of regulators on the applicability of the existing data protection regime to location data[4], it is not surprising that location data is now clearly referred to in the GDPR definition of personal data.

Online identifiers

The GDPR explains that online identifiers "may leave traces which, in particular when combined with unique identifiers and other information

[2] Efifiom Edem v The Information Commissioner and The Financial Services Authority [2014] EWCA Civ 92

[3] ICO guidance on Wi-Fi location analytics (February 2016)

[4] See for example ICO guidance on Wi-Fi location analytics (February 2016), Article 29 Working Party (A29WP) Opinions: WP115 (2005) WP185 (2011), WP216 (2014).

received by the servers, may be used to create profiles of the individuals and identify them"[5]. Examples include cookies and IP addresses.

The GDPR suggests that online identifiers of themselves will not always be personal data. However, given the multiplicity of data capture points in an online environment, it is unlikely that an entity (whether website operator, marketer, social media platform or other) will only ever be collecting one piece of information on a user. Combining information makes it far more likely that an individual will be identifiable, as many have already recognised for a while.

Does the GDPR really introduce 'new' categories of personal data?

Location data, IP addresses and MAC addresses

Location data, online identifiers and other identifiers are not explicitly mentioned in the Directive. However, the possibility that they may allow for the identification of individuals has been discussed in Europe for some time. The A29WP had already expressed the opinion that personal data could encompass IP addresses in 2007[6]. More recently, the ICO stated that IP addresses can be personal data, although usually only if the organisation handling it can access other information linked to that address that identifies an individual[7]. The ICO has also commented that "using a MAC address or other unique identifier to track a device with the purpose to single them out or treat them differently (e.g. by offering specific products, services or content) will involve the processing of personal data"[8].

Further guidance on the status of IP addresses should be provided by the Court of Justice of the European Union ('CJEU') in the coming months following the referral in the Breyer Case[9]. In particular, the CJEU has been asked to consider whether a dynamic IP address which a service provider stores when his website is accessed already constitutes personal data for the service provider if a third party (an access provider) has the additional knowledge required in order to identify the data subject. The opinion of the Advocate General ('AG') was that such IP addresses would be personal data. The AG's opinions are not binding on the CJEU but are often followed in practice. Given the direction of travel of recent CJEU decisions in strengthening individuals' right, this result would not be unsurprising.

Cookies

As for cookies and browser-generated information ('BGI'), the Court of Appeal held in 2014[10] that there was a serious issue to be tried as to whether BGI is personal data. BGI is tracking information processed via cookies by a search engine (Google in this case) to track the browsing activities of its users. The Court of Appeal looked at whether the BGI "individuates" the individual, "in the sense that they are singled out and distinguished from all others"[11]. The concept of 'singling out' is also mentioned in Recital 26 of the GDPR (see table above) but without further

---

[5] Recital 30 of the GDPR

[6] A29WP Opinion WP136 (2007)

[7] See ICO decision notice FER0529617

[8] ICO guidance on Wi-Fi location analytics (February 2016)

[9] Case C-582/14 Breyer v Bundesrepublik Deutschland

[10] Google Inc v Vidal-Hall [2015]

[11] An appeal to this decision (which only concerned an interim application to obtain permission to service proceedings outside the jurisdiction), although not on this point, has recently been withdrawn.

explanation. Further guidance from EU regulators on the exact meaning of singling out, when it might apply and how this concept interacts with measures such as anonymisation and pseudonymisation will be welcomed by many data controllers.

When does information relate to an individual?

The GDPR retains the terminology used in the Directive around information 'relating' to an individual. The recitals do nothing to clarify the meaning of this word, but many will be aware of the series of UK cases on this question (e.g. Durant v FSA), which continue to have relevance in the UK. It will be interesting to see whether controllers will seek to follow the bolder 'Durant' approach on biographical significance (particularly in the context of subject access requests) over the next few years. Brexit will likely make it easier for UK based data controllers to continue with this line of argument (unless the UK joins the EEA, in which case the GDPR will apply in full and a more harmonised approach to the word 'relate' would be expected across the EU).

Impact of the GDPR definition changes in practice

All organisations to which the GDPR will apply[12] should be checking their understanding of the types of data they process. For some types, such as location data, there are a number of entities that will process it - from network providers, to app developers, device manufacturers, social platforms, data platforms and standardisation bodies. Given the significant compliance burden the GDPR introduces, any incidental collection of location data should be revisited and, if possible, eliminated or modified (e.g. through anonymisation or pseudonymisation, discussed below). This will either place the data outside the scope of the GDPR or significantly reduce the risk of harm to individuals (and of enforcement action against businesses).

As for those organisations such as location-targeted advertisers who rely on the data they collect being personal data, they will clearly need to assess the impact of the GDPR on their business sooner rather than later. The extent of changes required will depend in part on how deeply a 'data hygiene' culture has been embedded throughout the business.

In general terms, organisations in the UK that were following the guidance of the A29WP and national regulators will already have been aware of the potential for location data, online identifiers and other identifiers to lead to the identification of individuals, especially when used in combination with other information. In this context, the GDPR is not a significant game changer. Indeed, one can expect the A29WP's successor, the European Data Protection Board, to draw on some of its more recent opinions as the basis of its guidance on the GDPR.

However, for those that were not closely aligned with such views and guidance, the penalties for ignoring them will now be much more severe. Maximum fines under the GDPR will be up to 4% of annual worldwide turnover.

[12] See Slaughter and May's briefing: New rules, wider reach: the extra-territorial scope of the GDPR

Can anonymisation and pseudonymisation help?

Anonymisation

Anonymisation was neatly defined by the ICO as "the process of turning data into a form which does not identify individuals and where identification is not likely to take place", resulting in data that is, therefore, not personal data[13]. Recital 26 of the GDPR echoes this by stating that "the principles of data protection should therefore not apply to anonymous information, that is […] data rendered anonymous in such a way that the data subject is not or no longer identifiable".

Crucially, the word 'likely' is missing from the GDPR definition. This implies a higher threshold than that put forward by the ICO. The ICO was of the view that 100% anonymisation, whilst desirable, is not required by the Data Protection Act 1998. Rather, the risk of re-identification needed only to be remote. The A29WP, on the other hand, argued that anonymisation must be irreversible[14]. Given that the volume and variety of personal data available in an online and connected world and the increased use of data analytics make re-identification more likely than ever, one would be forgiven for querying the usefulness of strict anonymisation as an effective tool for organisations to eliminate their compliance burden.

The GDPR does, however, make a small concession: when taking account of "all the means reasonably likely to be used" to identify someone, you can take account of all objective factors such as cost, available technology and amount of time. However, the time spent by organisations assessing these factors on a case by case basis may still mean that compliance with the GDPR requirements is the cheaper and quicker solution.

Pseudonymisation

Pseudonymisation is, in essence, a security measure organisations can apply to personal data, much like encryption. The GDPR explains that pseudonymisation, which was not explicitly mentioned in the Directive, is "the processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person".

What of hashing – does it amount to anonymisation? The ICO describes a hash function as a one-way method of converting information into a hashed value, often simply called the 'hash'[15]. When a user first provides information, for example credit card details, this is hashed and only this hash value is stored. When a user returns and enters their details again, the hash is freshly calculated then compared with the stored hash. If the two hashes match, then the user can be authenticated. The one-way nature of hashes is key: if an attacker somehow obtains a list of hashes, they cannot directly work out what the credit card details are, even if they know the particular hash function that was used.

However, there are other indirect methods of gaining access to such data and the ICO is quite clear that hashing of certain information such as passwords is unlikely to be enough to keep the

[13] ICO guidance on Anonymisation (2012)

[14] Opinion WP216 (2014)

[15] ICO guidance "Protecting personal data in online services: learning from the mistakes of others" (May 2014)

password safe. So if hashing carries a risk of re-identification, it will presumably be limited to being a form of pseudonymisation rather than effective anonymisation.

Despite the inherent and potential limitations of anonymisation and pseudonymisation, they should not be dismissed, if anything because they will help limit an organisation's risk profile and exposure in the event of a personal data breach. The GDPR specifically refers to pseudonymisation as one of the security measures that organisations should implement by default or "as soon as possible". In fact, given how often the GDPR refers to pseudonymisation, including as a 'safeguard', organisations should be giving serious attention to the possible uses of this process throughout their business.

Practical steps to prepare for the GDPR

Organisations should:

- consider running an information audit to verify their understanding of what personal data they collect and use;
- assess whether any data they process falls within the 'new' categories of personal data;
- ensure that where online identifiers identify an individual or make them identifiable, they are, at the very least, subject to the same safeguards as the 'conventional' personal data;
- consider introducing or expanding the use of pseudonymisation throughout the business;
- consider whether anonymisation techniques may be of assistance and whether they would be reversible in any given scenario. A data privacy impact assessment (see the ICO's 2014 guidance on Privacy Impact Assessments) is a useful tool to help with this assessment;
- monitor guidance from the A29WP and the ICO. Anything produced by them between now and May 2018 should take account of the GDPR requirements;
- check to ensure that contracts with third parties adopt the expanded definition going forward; and
- document as much of the above as possible. Accountability is a key concept in the GDPR and will require businesses to be able to demonstrate their compliance.

If you have any queries on this Briefing or if you would like to discuss any aspect of the GDPR or any data protection or privacy issue, please do not hesitate to contact Rob Sumroy, Rebecca Cousin or your usual Slaughter and May advisor.

Rob Sumroy
T +44 (0)20 7090 4032
E rob.sumroy@slaughterandmay.com

Rebecca Cousin
T +44 (0)20 7090 3049
E rebecca.cousin@slaughterandmay.com

Date July 2016