Big Data – an extract from contemporary views on GDPR

In HMG and MOD particularly, we used concepts of aggregation to determine our views on how to handle data. We looked at large amounts of data where each record on their own was not particularly sensitive, but given that a large data loss might allow an unlawful recipient to search that data the outcome, or consequence might actually become sensitive. If the search was by association corelating a person with a location and could then be 'married' to other sensitive information about the person identified the impact was thought potentially higher. We needed to treat that database not as SECRET but with the same due care and attention as if it was secret. Now with big data, aggregated data and the data science of analytics these associational qualities can be 'married' for purposes that do not serve the stated intention for the collection of each record. We need to think about Big Data in the sense of how it can be exploited, 'wrongfully' used for profiling by association to provide marketing information. We need to think how the rights of the citizen can be satisfied where back-ups are "somewhere in the cloud", on Iron Mountain tapes. Can this data be retrieved? Can we be sure that when we delete information on data subject request, we delete all copies and that it cannot reappear through the magic of metadata manipulation? Can we be sure that anonymization or pseudonymising or encryption really work and cannot be 'undone'?

Big data is "all about seeing and understanding the relation within and among pieces of information that, until very recently, we struggled to fully grasp." Discovering these new relationships is the work of analytics — the automated processing of data. This central feature of big data faces differing perceptions that produce ambivalence in the General Data Protection Regulation that will affect how big data is used.

On one hand, the European Commission is placing a big bet on big data in its strategy for economic growth. The EU's 2015 Digital Single Market Strategy targets big data as "central to the EU's competitiveness" and a "catalyst for economic growth, innovation and digitisation across all economic sectors [...] and for society as a whole." On the other hand, use of automated processing — algorithms — touches a deep vein of distrust of computing and data use. As put by the Article 29 Working Party (WP29), "the real value of big data still remains to be proven."

These differing views are reflected in the debates on rights concerning "profiling" in Article 22 of the GDPR, which addresses automated decision-making and profiling, and in related provisions. The end result is a compromise provision that was one of the last issues to be resolved in the GDPR Trilogue. It provides, "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Article 22 is one of several opaque provisions that leave the full effect of the GDPR on analytics and big data open to interpretation and need to be resolved in the implementation and interpretation of the GDPR before May 2018.

The Legislative debate on profiling

"Profiling" — an algorithmic inference drawn from data about an individual — is a primary tool for extracting value from big data. It is used widely in legitimate ways that can be helpful to society and individuals, and a provision applying to all profiling without differentiation between benign profiling and profiling that can cause harm to individuals could diminish beneficial uses of analytics and impede the commission's aspirations for big data in EU digital development. On the other hand, certain uses of algorithms can present risks. Quoting from the 2014 White House Big Data Report, the EDPS identified risks of big data as lack of transparency, informational imbalance, erosion of data protection principles, spurious correlations and unfair discriminatory outcomes, as well as social, cultural, and creative stultification.

As originally proposed by the commission in 2012, the profiling provisions (then Article 20(1)) would have applied to a measure based solely on automated processing which "produces legal effects ... or significantly affects" him or her. Users of analytics raised concerns that the vagueness and breadth of this language could have unintended consequences for many widely-adopted business practices.

The WP29 agreed this "significantly affects" test was "imprecise," and concluded that the commission's language was insufficient to protect individuals in the creation and use of profiles. The European Parliament appeared to respond to these concerns by adopting an express right to object to profiling based solely or predominantly on automated processing where such profiling "leads to measures producing legal effects ... or does similarly significantly affect" an individual. Although, the Parliament's language was more precise than the Commission's proposal, it was wide enough to include creation of a profile as such, and not just the use and effect of a profile. At the same time, the Parliament added provisions to encourage use of pseudonymised data (data "which does not permit the direct identification of a natural person but allows singling out of a data subject"); these included Recital 58(a) providing that profiling based on pseudonymised data "should be presumed not to significantly affect the interests, rights or freedoms of the data subject."

The European Council debated Article 20 at length. Industry proposed an alternative test based on automated processing that is "unfair and discriminatory" to a data subject; other proposals in essence would have introduced a harms threshold. The Irish Presidency in particular suggested a test of "severe negative effects," and this language was actively considered during the subsequent Lithuanian and Greek Presidencies. There was insufficient consensus, however, and ultimately the Council reverted to substantially the test proposed by the Commission (i.e. "significantly affects"), but amended it to focus on the effect on the individual rather than either profiling as such or the "measures" used for profiling.

The compromise outcome in Article 22

In the end, the Council's focus on effects (a "decision" rather than simply a "measure"), combined with compromise wording proposed by Parliamentary negotiators (a decision which "produces legal effects concerning [an individual] or similarly

significantly affects [an individual]"), was the basis for agreement in the Trilogue and finally adopted as Article 22. Read together with other provisions of the GDPR affecting profiling and automated processing, the essence of the right not to be subject to certain automated decisions is not to foreclose altogether the use of automated processing but to require ways for humans beings to become involved in such decisions. Recital 71 gives as examples "refusal of an online credit application or e-recruiting practices" where a decision is "automatic" or "without human intervention."

This reading of Article 22 is reinforced by other provisions addressing profiling. For example, Articles 13 and 14 require businesses engaged in profiling to provide individuals with "meaningful information about the logic involved [in the profiling], as well as the significance and the envisaged consequences of such [profiling] for the [individual]." Similarly, Article 15 requires a business to disclose "the existence of automated decision-making, including profiling" upon request. These provisions enable a basis to invoke the human intervention required by Article 22.

Recital 63 limits compliance with a request for access to information about profiling to where disclosure does not adversely affect the rights of others, "including trade secrets or intellectual property and in particular the copyright protecting the software." This suggests that the disclosure required by Article 15 involves the fact and general logic of profiling and decision-making, but not the specifics of algorithms that implement this logic.

These requirements may necessitate revisions to use specifications and similar content in privacy policies and other customer information. More significantly, many organizations will have to put in place mechanisms to respond to requests for access to profiling information and enable review of decisions subject to Article 22.

Additional provisions reflect the view that profiling poses inherent risks for individuals. Article 35(3)(a) requires privacy impact assessments whenever a business is engaged in "systematic and extensive processing" subject to Article 22. This provision effectively deems such processing to be per se in the category of processing "likely to result in a high risk to the rights and freedoms of natural persons" to which Article 35 applies. Article 21(2) extends the right to object to profiling "to the extent it is related to direct marketing" or based on legitimate interest or public interest unless the controller "demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject." Article 21(4) requires special notice of this right "explicitly brought to the attention of the data subject … clearly and separately from any other information." Recital 70 makes clear this right to object applies "at any time and free of charge." This right is likely to spawn new processes and procedures and potentially changes to business systems.

The GDPR contains two sets of provisions that ease the Article 22(1) restriction on profiling. First, Article 22(2) contains three specific exemptions. These include where a decision based on automated processing is: (i) necessary for entering into or performing a contract; (ii) authorized by EU or Member State law, or (iii) based on the explicit consent of the individual. Even where these exemptions apply, however, "suitable measures" must safeguard the interests of the individual including, as a minimum, the right to obtain human intervention for the individual to express his or her

point of view and contest the decision involved (Article 22(3)). In addition, Recital 71 emphasizes that where these exemptions apply, the processing must be fair and transparent; statistical procedures adequate, and appropriate measures implemented to ensure security, prevent discriminatory effects, and minimize the risk of errors and ensure inaccuracies are corrected. Further, profiling should not involve children.

Second, the GDPR allows for certain uses of pseudonymisation. The Trilogue removed the exemption proposed by Parliament as Recital 58(a), but inserted numerous other provisions throughout the GDPR where the use of pseudonymous data is encouraged as a means of implementing appropriate safeguards. These include: (i) Article 6(4) (pseudonymisation a factor controllers should consider when determining compatibility of purpose for further processing); (ii) Article 32(1), (pseudonymisation may assist controllers in meeting security requirements); (iii) Article 25(1) (pseudonymisation an example of a measure that may satisfy requirements for privacy by design); (iv) Article 40(2) (encouraging adoption of codes of conduct that promote use of pseudonymization), and (v) Article 89(1) (pseudonymisation as a means to data minimization when processing personal data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes). Pseudonymisation, therefore, can be used for "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests" as required under Article 22(3) where an organization wants to rely on one of the Article 22(2) exemptions. (Note that pseudonymous data will be treated as personal data under the GDPR, and therefore must be processed in accordance with its requirements.)

What Article 22 means for big data in practice

Article 22 and its companions present challenges for analytics and other automated processing that uses personal data as an input. The substantive limits as well as procedural checks erect significant speed bumps on the road toward deployment of big data under the EU's Digital Single Market Strategy.

The size and frequency of these speed bumps will depend on interpretation, guidance, and application as the GDPR is implemented. Article 70 tasks the European Data Protection Board (EDPB) with issuing guidelines, recommendations and best practices for consistent application of the regulation and, in particular, "further specifying the criteria and conditions for decisions based on profiling pursuant to Article 22(2)".

A key threshold question is whether the right "not to be subject" to automated decision-making based on profiling is satisfied where a business implements measures to safeguard individuals' interests such as a way for individuals to contest a decision with legal or similarly significant effects, for example. As discussed above, the language in Recital 71 on "automated" decisions and "human intervention" suggests this is the case, but the language of Article 22 itself is not clear. If this is not the case, then Article 22 will restrict numerous well-established and effective practices.

A second key question is when the impact of an automated decision amounts to "legal effects" or "similarly significant" effects. Recital 70 gives the examples of denial of credit or e-hiring. Insurance coverage and perhaps rating decisions likely fall into the same category (but might be considered necessary for entering into a contract for the purposes of Article 22(2)). In contrast, many other uses of profiling are innocuous (though some might include marketing in this category notwithstanding its targeting in Article 21(2)), but there is a spectrum of uses that will require judgments as to their effects. The language of Article 22(3) and (4) suggests that covered effects must be enough to touch a data subject's "rights and freedoms and legitimate interests." Similarly, the Parliament framed legal effects and similarly significant effects in terms of effects on "the interests, rights or freedoms of the concerned data subject."

Other questions include the definition of profiling, the circumstances in which exemptions in Article 22(2) can be relied upon, when automated decisions can use sensitive personal data, what constitutes "compelling legitimate grounds" permitting a controller to disregard an objection to profiling, and what degree of human intervention would be required to bring automated processing (including profiling) outside the restriction in Article 22(1).

The EDPB also will provide guidance on privacy impact assessments, a subject included in the WP29 work plan for 2016. This overlaps with Article 22 because of the requirement in Article 35(3)(a) to conduct privacy impact assessments for certain automated processing. This new requirement may be a significant burden for many businesses, in particular where they conclude that significant risks cannot be mitigated and therefore are obliged to consult with the data protection authority or seek the views of the data subject (where appropriate).

The "explicit consent" exemption in Article 22(2)(c) is likely to be challenging to apply to numerous profiling activities. In addition, where automated decisions rely on sensitive personal data, explicit consent may be the only legal basis for processing unless the profiling is authorised by EU or Member State law.

To prepare for application of Article 22, businesses should review current profiling and automated processing and consider: (1)

- Do these result in decisions involving individuals?
- What are the effects of such decisions and do they affect legal interests of any individuals or other significant interests?
- What data are the decisions based on and does they employ sensitive data?
- What is the algorithmic logic involved and can it be explained to data subjects?
- Is there some element of human intervention in decisions based on the automated processing?

To the extent profiling and processing are subject to Article 22(1), organizations should determine if these activities are covered by one of the limited exemptions. Specifically, they should consider whether:

- If they want to rely on explicit consent of the data subject, the consent is valid (i.e. complies with requirements under Article 7 and Recitals 32 and 43);

- If they want to rely on the "contract exemption," in accordance with WP29 guidance the processing is genuinely necessary and not unilaterally imposed on the data subject by the controller;
- They have implemented suitable measures to safeguard the personal data (such as pseudonymisation);
- They have carried out a privacy impact assessment, and
- they have in place a process for data subjects to obtain human intervention. If the activities do not fall within exemptions, they should be modified.

How the EDPB carries out its assigned task to interpret Article 22 will have a significant impact on the future of big data in the EU as well as on existing business practices. Presumably with these impacts in mind, the Commission weighed in on the relationship of the GDPR to big data even before the legislative process was complete, stressing that the regulation "will offer flexibility to businesses all while protecting individuals' fundamental rights." It pointed to provisions promoting anonymisation, pseudonymisation, and encryption and stated that "[c]ompanies are free to base processing on a contract, on a law or, on, in the absence of other bases, on a 'balancing of interests.'" How the EDPB addresses these issues will decide whether Article 22 is as flexible as the Commission's communication suggests. As the WP29 establishes the EDPB and pursues consultations on interpretation of the GDPR, it should provide a robust opportunity for input from the Commission, industry, civil society, and other stakeholders — including written notice and comment — to ensure a full appreciation of the implications of the GDPR in application.