IPS208: Key GDPR Legal Issues regarding Privacy-Preserving Techniques

Dr W Kuan Hon[1*]

[1] UN Privacy Preserving Techniques Task Team; London, UK; unppttt@kuan0.com

Abstract:
This paper provides a GDPR tutorial for non-data protection practitioners, covering controllers, processors, processing, personal data, anonymisation, pseudonymisation, special category and criminal-related data; GDPR's core principles and other obligations, especially security, data protection by design and by default and data protection impact assessments; relaxations for statistical processing; and GDPR's interrelationship with privacy-preserving techniques (PPTs), notably input privacy, secure multi-party computation (MPC) and trusted execution environments (TEEs), focusing particularly on cloud computing and transfers under GDPR.

Keywords:
Data protection by design and by default; EU data protection laws; privacy-enhancing techniques; transfers; cloud computing.

1. Introduction and scope:

This paper explains, for statisticians and others who are not data protection practitioners, the relevance and importance of the European Union (EU) General Data Protection Regulation (GDPR), and GDPR's interrelationship with privacy-preserving techniques (PPTs), particularly regarding input privacy (based on the position as at 2 June 2021). There are other EU laws on data protection. Personal data processing by law enforcement authorities is regulated instead under the Directive (EU) 2016/680, while personal data processing by EU institutions, such as the European Parliament, Commission and Eurostat, is regulated under Regulation (EU) 2018/1725 (supervised by the European Data Protection Supervisor, EDPS). This paper discusses GDPR, but most of it is also relevant to processing under those other laws.

2. GDPR - overview:

The GDPR applied in the European Economic Area (EEA) from May 2018, significantly changing the region's data protection laws and enforcement. Post-Brexit, the UK's national laws retain the "UK GDPR", in substance currently broadly the same as the EU GDPR's. The GDPR aimed to modernise and harmonise EEA data protection laws, yet allows EEA Member States to take different approaches in certain areas. So, EEA data protection laws are still not fully harmonised, and *national* laws remain relevant. Basic familiarity with the GDPR is important, not just because of the potential risks discussed below, but because it motivated many other countries to enact similar laws. This, with GDPR's "gold standard" rules, means that compliance with GDPR will likely enable compliance with other data protection/privacy laws too, thus assisting multinational projects (though relevant local laws still need checking).

It is well-known that national supervisory authorities (SAs) charged with enforcing GDPR (e.g. the Netherlands' Autoriteit Persoonsgegevens) can impose administrative fines, whose ceiling depends on the rule infringed: the two tiers are 2% of turnover/(if higher) €10m, or 4% of turnover/(if higher) €20m. A lesser-known point is that EEA Member States can decide whether and to what extent their public authorities and bodies can be fined, if at all.

Hitherto, attention has focused more on fines than on "data subjects", living human beings, being able to claim compensation for physical, material or "non-material" damage if their

"personal data" is "processed" in a GDPR-infringing way (quotation marks are used as these terms have special meanings under GDPR, discussed below.) Total payouts could dwarf fines if quasi-class actions succeed, where representative organisations may claim, including through lawsuits, compensation on behalf of thousands, even millions, of affected data subjects (Art.80 GDPR). Such court litigation is increasing, often funded by organisations that plan to receive a share of any payouts. SAs can order suspension or termination of processing (higher-tier fine for disobeying orders) - which could also be disruptive and costly.

3. GDPR - controllers, processors, processing and scope:

The GDPR regulates the "processing" of "personal data" by "controllers" and "processors". It does not regulate *data* as such, but rather *actions* relating to personal data, i.e. "processing".

National SAs supervise and enforce its compliance, but the European Data Protection Board (EDPB), comprising SAs plus the European Commission and EDPS, publishes guidance on interpretation/application and resolves disputes between national SAs. EDPB and national SA guidance is authoritative, but strictly not legally binding. It is EU courts, going up to the Court of Justice (CJEU), that rule definitively on EU laws' meaning, including GDPR's.

"Processing" is a very broad GDPR concept: "any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" (Art.4(2). Basically, doing *anything* to or with electronic personal data is "processing", including storing/hosting, viewing, or transmitting it - not just using, crunching, disclosing or sharing it.

Controllers control the "purposes and means", the *why* and how, of processing personal data. They are primarily "on the hook" to ensure compliance with GDPR. (The GDPR also introduced a "joint controller" concept (Art.26), but discussion here is out of scope.)

Controllers may engage processors to help them process their personal data, e.g. using a hosting or cloud provider to store personal data. Processors may engage sub-processors (termed "another processor" under Art.28(2)), and so on. Processors have far fewer direct duties under the GDPR, notably regarding security, use of sub-processors and "transfers" (see below), but processors also have contractual obligations (and liabilities) to their controllers under the agreements GDPR requires to be in place between controllers and their processors. Controllers/processors may be organisations, public bodies or individuals.

Theoretically, compensation claims may be made against *any* actors in a personal data processing supply chain (controller, processor, sub-processor etc), e.g. because they are perceived to have "bigger pockets" or are closer locally; and they may have to pay out in full, even when the fault lies largely with some other actor(s) (Art.82(4)). The GDPR exempts from liability processors who have complied with all their controllers' lawful instructions and all GDPR's processor-directed obligations (Art.82(2)), and exempts those who prove they are not "in any way responsible" for the damage to data subjects (Art.82(3)). Full compensation paid can be claimed back from others actors according to their responsibility for the damage (Art.82(5)). However, Art.82 could be clearer, and it allows reimbursement only for *compensation* paid, not legal fees or other costs of proving lack of responsibility; so those against whom claims were initially made may still have to expend time and money to defend themselves, time and money they may not be able to claim back under GDPR. Therefore, legal experts need to be involved from the outset, to ensure contracts between supply chain actors fully cover indemnity and reimbursement issues.The GDPR has both a "material scope" and a "territorial scope". Certain processing is *exempt* from its material scope, e.g. processing for "purely personal or household" purposes (e.g. individuals' personal address books storing others' personal data), and processing for national security purposes. Even if within the GDPR's material scope, processing is caught by the GDPR, and must follow GDPR rules, *only* if it is *also* within the GDPR's territorial scope. Space and this paper's remit do not permit discussion of the GDPR's complex territorial scope issues. All processing discussed here is assumed within both material and territorial scope.

4. "Personal data", anonymisation and pseudonymisation under GDPR:

"Personal data" is a critical GDPR concept, because processing must comply with GDPR's requirements *only* if it involves "personal data". The GDPR does not apply when processing anonymous data or anonymised data. The GDPR's "personal data" concept is *far wider* than the US concept of "personally identifying information" (PII): it is "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier *[such as IP address or MAC address]* or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Art.4(1)).

To determine whether a natural 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 another person* to identify the natural person *directly or indirectly*. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the *costs of and the amount of time* required for identification, taking into consideration the *available technology* at the time of the processing and *technological developments*" (Rec.26). Accordingly, information is "personal data" if it can be combined, whether by the controller *or anyone else* (e.g. a member of the public), with other information that is reasonably likely to be obtainable from any sources, in order to identify or "single out" an individual (e.g. *distinguish them* or *treat them differently* from other individuals - even if their name/identity is not known).

This means that, in practice, it is very difficult to anonymise personal data effectively so that it can be processed without regard to the GDPR. Also, the *procedure of anonymising* personal data itself constitutes "processing", and must be conducted in compliance with GDPR's rules. Aggregation is probably the most effective method of anonymisation (Information Commissioner 2012), but bearing in mind the "means reasonably likely" test.

Under the GDPR, "pseudonymisation" also has a special meaning: "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person" (Art.4(5)). Key-coding pharmaceutical data and encryption are both forms of pseudonymisation. Note that the procedure of pseudonymising personal data is *itself* "processing"; also, *pseudonymised data is still "personal data"* to which GDPR applies (Rec.26). In the GDPR context, pseudonymisation is treated as a *safeguard* for personal data, to reduce risks to data subjects (e.g. Recs.28, 156, Art.6(4), and 7 below), but not as a way to fully anonymise personal data.

5. GDPR's core principles and other obligations:

Controllers must comply with certain core principles when processing personal data (Art.5): "lawfulness, fairness and transparency"; "purpose limitation" (processing personal data *only* for the purposes for which it was collected or "compatible" purposes); "data minimisation" (throughout the processing lifecycle); "accuracy"; "storage limitation"; "integrity and confidentiality"; and "accountability" (higher-tier fine for infringement). Those terms again have special meanings and nuances under the GDPR, not always corresponding with their ordinary language meanings. Some principles also overlap with each other (e.g., arguably storage limitation is a subset of data minimisation), or with other GDPR obligations (e.g., integrity and confidentiality with security; and data minimisation of course assists confidentiality).

Furthermore, controllers are not permitted to process personal data without a recognised legal basis or lawful basis under GDPR, i.e. meeting one of the conditions or gateways set out in Art.6(1). Data subject consent is one gateway, but *not* the only one. Because consent is so difficult to obtain validly under GDPR (see Art.7's conditions), other legal bases are used much more often, e.g. the processing is necessary to perform a task in the public interest or to exercise official authority vested in the controller (such as national official statistics bodies and other public bodies); or the processing is necessary in the legitimate interests of the controller or a third party, where are not outweighed by data subjects' interests, rights and freedoms.

Special category data (personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation) cannot generally be processed unless one of certain conditions is met, e.g. the data subject's *explicit* consent to *specified processing purpose(s)* (Art.9). Personal data relating to criminal convictions and offences can be processed *only* under the control of official authority, or when the processing is authorised by EU or Member State law providing appropriate safeguards for data subjects' rights and freedoms (Art.10). These restrictions are *in addition to* the usual requirement for there to be a legal basis for processing that data, e.g. public task, and other GDPR rules.

Controllers must comply when individuals exercise their data subject rights, i.e. rights in certain situations to access, correct or erase (a.k.a. the right to be forgotten) their personal data, restrict its processing, "port" it elsewhere, and/or object to certain processing (higher-tier fines). GDPR also restricts controllers' "automated decision-making" (higher-tier fine).

Other obligations are imposed on controllers, e.g. issuing privacy notices, and on processors. When using processors to host or otherwise process personal data, controllers must conduct due diligence on the processors' ability to maintain GDPR compliance, ensure its processor contract contains certain mandatory minimum terms, and conduct regular post-contract audits/checks on the processor's compliance with GDPR and the contract, while processors must "flow down" these contract terms to their sub-processors, etc. (Art.28).

Other rules (Ch.V) restrict "transfer" of personal data to so-called "third countries" outside the EEA (or UK), or to international organisations (e.g. the United Nations), unless certain mechanisms are used (higher-tier fine for controllers *and processors*). "Transfer" here means both data exports or international transfers, i.e. transmitting or physically hosting personal data outside the EEA, e.g. for backups, *and* allowing an organisation or individual *outside* the EEA *remote* access to EEA-hosted personal data, e.g. for maintenance or support (EDPB 2020). The GDPR's transfers restriction, also considered to be a law requiring data localisation (or "data sovereignty", which term has no real meaning legally), has particular implications for cloud computing (Hon 2017).

6. Security, data protection by design and by default and data protection impact assessments under GDPR:

We now turn to certain GDPR obligations that are key in the context of PPTs. Art.5(1)(f) requires "integrity and confidentiality, while Art.32 further requires implementation of appropriate technical and organisational measures to ensure a level of security appropriate to the risk to data subjects. The main differences: "integrity and confidentiality" applies only to controllers (higher-tier fine); but Art.32 applies to controllers *and processors* (lower-tier fine).

Controllers must implement "data protection by design and by default" (DPbDD, Art.25). Similar, but not identical, to the broad "privacy by design" concept, DPbDD requires controllers, taking into account the state of the art (available technology), costs, the nature, scope, context and purposes of the processing and risks to individuals' rights and freedoms, to implement appropriate technical and organisational measures (e.g. pseudonymisation), *designed* to implement data-protection principles (e.g. data minimisation), in an effective manner and to integrate the necessary processing safeguards, in order to meet the GDPR's requirements and protect data subjects' rights. This applies *both* when determining the processing means (e.g. what techniques to use), *and* when conducting the processing itself.

Controllers must also implement measures to ensure that, *by default*, only personal data necessary for each specific processing purpose is processed - including how much data is collected, the extent of their processing, their storage period and their accessibility (e.g. least privilege access controls). Measures must also ensure that *by default* personal data are "not made accessible without the individual's intervention to an indefinite number of natural persons" (e.g. published, or openly exposed on the Internet).

Thus, DPbDD covers more than just security, though all techniques to protect security (whether confidentiality, integrity or availability), as "security by design and by default", are a subset of DPbDD. Security techniques are not discussed further here.

Data protection impact assessments (DPIAs) under Art.35 merit particular mention. Before commencing any processing that poses "high risk" to data subjects, controllers must assess its impact on data protection. DPIAs must *always* be conducted in certain situations, including "large scale" processing of special category or criminal-related data. Also, SAs can produce their own lists of processing activities that require (or conversely do *not* require) DPIAs, so again there may be national differences. DPIAs must contain certain minimum content, including measures e.g. safeguards to eliminate or mitigate the risks assessed. If despite those measures high risks to data subjects remain, the controller cannot commence the processing without first consulting the SA, who can approve it - or prohibit it altogether (Art.36). DPIAs are sometimes published by some controllers, particularly public bodies, whose DPIAs may also be obtained under freedom of information or access to information requests.

7. Relaxations under GPDR - processing for statistical purposes:

GDPR allows various relaxations in relation to processing for statistical purposes (and for certain archiving and research purposes - but this paper discusses only processing for statistical purposes, here termed "statistical processing").

Personal data may be processed for statistical purposes compatibly with "purpose limitation", and may be stored for longer periods despite "storage limitation", *provided* appropriate technical and organisational safeguards for data subjects are implemented under Art.89 (particularly data minimisation, including by processing only anonymised or pseudonymised data) (Art.5). Special category data may be processed if necessary for statistical purposes, with safeguards, based on EU or Member State law that is proportionate, respects data protection rights and provides suitable specific safeguards. For statistical processing, in some situations, privacy notices are unnecessary and erasure and objection rights may not apply. Member States may, with safeguards, exempt statistical processing from *other* data subject rights that would make such processing impossible or seriously impair it. However, no other exemptions exist for statistical processing, e.g. from the transfer restriction.

8. Input privacy, privacy-preserving techniques, MPC and TEEs under GDPR:

Input privacy has a narrow meaning, i.e. that a party computing on data provided by third parties cannot access or derive any provided input value or access intermediate values or statistical results during its processing, unless the value has been specifically selected for disclosure; with PPTs for input privacy including secure multi-party computation (MPC), zero knowledge proofs and trusted execution environments (TEEs) (Craddock et al. 2019).

Such PPTs are still developing, but MPC and TEE services are now publicly available, and increasingly efficient. MPC involves jointly computing an agreed-upon function among a set of possibly mutually distrusting parties while preventing any participant from learning anything about inputs provided by other parties, and TEEs involve the use of special-purpose hardware and software enabling a process to run without memory or execution state being visible to any other process, not even the operating system or other privileged code (Craddock et al. 2019).

Mapping concepts of PPTs or "input privacy" to the GDPR is not straightforward. Those terms generally relate to confidentiality but, as noted at 6 above, GDPR's concepts of security and DPbDD extend beyond confidentiality-preserving techniques to techniques for compliance with its multifarious requirements, particularly data minimisation measures. However, in this context, input privacy is most relevant to *controllers' use of processors* in situations where *only controllers* should know the inputs and outputs while processors should not - particularly in relation to security, DPbDD, as risk-reducing safeguards to be considered under DPIAs, and to enable transfers to processors outside the EEA (or UK, for the UK GDPR).

In July 2020, what's popularly known as the *Schrems II* case (C-311/18, *Data Protection Commissioner v Facebook Ireland Limited and Maximillian Schrems* ECLI:EU:C:2020:559) invalidated the EU-US Privacy Shield for transfers to the USA, but allowed so-called standard contractual clauses between transferor and transferee (SCCs) to continue to be used to enable transfers to third countries generally, if "supplementary measures" to protect personal data are implemented. Helpfully, draft regulatory recommendations cite implementation of MPC, "split processing" using processors in different countries, as an acceptable technical

supplementary measure for transfers outside the EEA (or UK) to such processors, and also a "by design" organisational supplementary measure for data minimisation to prevent unauthorised access (EDPB, 2020). Pre-GDPR, use of certain MPC even in public cloud was considered by Estonia's SA to involve processing of anonymised data (LaGrone 2017).

However, the draft recommendations failed to mention TEEs altogether, despite their now-wider availability and practicability, including in cloud computing (e.g. AWS (Barr 2018), Google (Porter et al. 2018), and Microsoft Azure (Russinovich 2017), also see Hon (2020)). Use of TEEs may be less efficient and affordable (and also much less available) with SaaS services, as compared with IaaS/PaaS services. However, omitting discussion of TEEs seems inexplicable: their use can give controllers confidence to use processors, cloud or otherwise, in the EEA/UK or outside, without processors gaining knowledge of the personal data or processing operations, thereby also reassuring the controllers' SAs regarding security and DPbDD as well as the adequate protection of transferred personal data.

9. Summary and Concluding Remarks:

PPTs can significantly assist compliance with GDPR and other data protection/privacy laws. Experts in GDPR and relevant local laws should be involved when assessing their use, in view of such laws' complexity and risks. MPC and TEEs in particular have much potential, notably in the context of cloud computing and also "transfers". The EDPB, SAs and other regulators should be encouraged to recognise specifically the role of TEEs and other "confidential computing" techniques to enable compliance not just with the GDPR's transfers restriction but much more generally, including for security and DPbDD. PPTs should also be considered when conducting DPIAs, as possible safeguards to reduce the risks to data subjects below "high", thereby also obviating the need for prior SA approval to the intended processing.

References:

1. Barr, J. (2017). Amazon EC2 Update - Additional Instance Types, Nitro System, and CPU Options. https://aws.amazon.com/blogs/aws/amazon-ec2-update-additional-instance-types-nitro-system-and-cpu-options/ 19 June 2017; accessed 2 June 2021.
2. Craddock et.al. (2019). UN Handbook on Privacy-Preserving Techniques. UN. https://docs.google.com/document/d/1GYu6UJI81jR8LgooXVDsYk1s6FlM-SbOvo3oLHglFhY/edit 12 March 2019; accessed 2 June 2021.
3. EDPB (2020). Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data.
4. Hon, W. K. (2017). *Data localization laws and policy - the EU data protection international transfers restriction through a cloud computing lens*. Edward Elgar.
5. Hon, W. K. (2020). Schrems II additional safeguards: confidential computing. https://blog.kuan0.com/2020/08/schrems-ii-additional-safeguards.html 17 August 2020; accessed 2 June 2021.
6. Information Commissioner (2012). Anonymisation: managing data protection risk code of practice. https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf; accessed 2 June 2021.
7. LaGrone, C. (2017). National Special Education Data Analysed Securely. https://sharemind.cyber.ee/national-special-education-data-analysed-securely/, 7 December 2017; accessed 2 June 2021.
8. Porter et al. (2018). Introducing Asylo: an open-source framework for confidential computing, https://cloud.google.com/blog/products/identity-security/introducing-asylo-an-open-source-framework-for-confidential-computing 3 May 2018; accessed 2 June 2021.
9. Russinovich, M. (2017). Introducing Azure confidential computing. https://azure.microsoft.com/en-us/blog/introducing-azure-confidential-computing/ 14 September 2017; accessed 2 June 2021.