Processing pseudonymised data for research, statistics and archiving: Korea in the shoes of an EU Member State

Graham Greenleaf, Professor of Law & Information Systems, UNSW Australia

DRAFT 13 January 2020

AI, Ethics & Data Governance - Workshop on January 20, 2020, Korea University Seoul, Korea

Background provided by the organisers

Background provided: 'What is the scope of permitted "further processing" for the new purposes "compatible" with original purposes? What is the scope of public interest archiving, scientific/historical research, and statistics for which personal data can be used for new purposes? What is the required features of pseudonymization that are required for such further processing?'

Background provided: ' 'GDPR allows non-consensual 'further processing' of data for new purposes compatible with original purposes where public-interest archiving, scientific/historical research, and statistics are deemed to be compatible as long as safeguards are in place where the only explicitly stated safeguards are pseudonymization. These provisions were together interpreted by many that pseudonymized data can be used for scientific research not included in the original purpose.'

Background provided: ' 'In an attempt to emulate GDPR, Korea's Personal Information Protection Act amendment, which now passed the relevant congressional committee, states that in the newly inserted Article 28-2 that "data processors may process pseudonymized data for statistics, scientific research, and public interest archiving, etc ., without the consent of data subjects." where "Scientific research" means "development and verification of technologies, basic research, applied research, and private sector invested research, and other research using scientific methods."

Approach taken in presentation

The organisers have requested that I comment on six questions.

A reliable English translation of the whole Korean PIPA, as proposed to be amended, is not available. Others are therefore better placed to discuss what the revised PIPA might mean in relation to these questions.

I will therefore approach the questions as if Korea was an EU member state, and attempt to answer these questions in terms of what the GDPR requires or allows. However, I stress that in doing so I am not purporting to answer the question 'what will the EU require in Korean law in order for Korea's protections to be regarded as adequate under GDPR art. 45?'

Accurate prediction of what implications any of the PIPA changes might have for an EU assessment of the 'adequacy' of Korea's protections is not possible. Under the GDPR (as with the Directive), an 'adequate level of protections' does not require identical protections, although the CJEU in *Schrems I* held (under the Directive) that they must be 'essentially equivalent'. The EC's decision concerning Japan (2019) did not clarify what this meant.

Applying the GDPR

The following analysis applies the GDPR to the use of personal data for archiving, research, and statistics ('ARS uses'), and the use of pseudonymous data, from the perspective of a EU Member State. It addresses the six questions, but not necessary in the order given.

1.1. Pseudonymous data is still personal data, and not *per se* exempt

Pseudonymous data is personal data for GDPR purposes: 'The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person' (Recital 26). See also art, 4(5) definition of 'pseudonymisation'.

In contrast, processing of anonymous information 'including for statistical or research purposes', does not come within the GDPR. 'Anonymous information' is described as 'information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable' with 'identifiability' further described[1] (Recital 26)

Secondary uses of pseudonymous data are not *per se* exempt from the art. 6(4) 'compatible uses' test – only certain uses are exempt. Pseudonymisation is important in this discussion, not as a basis for further processing, but only as a possible means of satisfying the need for safeguards.

1.2. Only four categories of data uses are presumed to be compatible

The GDPR deems as satisfying the 'compatible uses' test only four categories of uses of personal data: 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' ('ARS uses'). For these uses, the 'presumption of compatibility' applies. (Korea has not included 'historical research' in new art. 28-2, so it only covers three of these categories.)

There is no 'etc' in the list of categories of uses deemed compatible. The GDPR consistently refers in eleven articles and recitals[2] to 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes', which is a closed list. There is no explicit additional expression such as 'etc' or 'and similar purposes' to convert this into an open list. Nor is there any justification for an argument that it is implied that similar purposes are also excepted.

Only 'ARS uses of pseudonymised data' are exempt from the art. 6(4) criteria for compatibility Any uses of pseudonymised data other than ARS uses are not presumed to be be compatible.

Any secondary uses of pseudonymised data to for other than RSA uses (in particular, to make decisions affecting individuals) are not presumed to be compatible. Any such uses must therefore satisfy the art. 6(4) compatibility test, taking into account factors (a)-(e) listed therein (link between purposes; context of collection; nature of the personal data, including whether special/sensitive data; possible consequences of processing; appropriate safeguards). Many proposed uses are likely to fail such a test. Of course, the use of personal data to make

---

[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. [4]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.' (GDPR Recital 26)

[2] GDPR Recitals 50, 52, 53, 62, 65, 156 ; Arts. 5(1)(b), 5(1))e), 9(2)(j), 17(3)(b), 89

decisions about individuals will often be the *primary* purpose of collection of the data, in which case it must have had a lawful basis for processing under art. 6(1).

1.3. The rest of the GDPR still apples to ARS uses of pseudonymised data

Although the 'presumption of compatibility' applies to ARS uses of pseudonymised data, since it it still personal data, the starting point must be that the rest of the GDPR still applies, unless the GDPR provides otherwise.

Four provisions may limit the extent to which the rest of the GDPR applies to ARS uses:

(1) Does art. 6(1) requiring a lawful basis for processing, apply even to secondary ARS uses? Where data is used for *compatible* secondary purposes (in accordance with art. 6(4)), Recital 50 provides that 'no legal basis separate from that which allowed the collection of the personal data is required', and further that ARS uses 'should be considered to be compatible lawful processing operations'. It seems logical that, since ARS uses are *presumed to be compatible* with the purpose of collection, no separate lawful ground for processing is needed, and Recital 50 implies this. However this conclusion is not explicit or free from doubt,[3] and if this is not so, then pseudonymised data cannot automatically be used for ARS purposes, and this use has to have a separate lawful ground, based on art. 6(1). This will often be readily found (for example in (a) consent or (f) legitimate interests). Recital 33 envisages that broad consent can be given for future scientific research purposes.

(2) The duration of storage of personal data 'in a form which permits identification of data subjects' (which includes pseudonymised data) is limited by art. 5(1)(e), but storage 'for longer periods [is allowed] insofar as the personal data will be processed *solely*' for ARS uses (emphasis added). In other words, longer storage for other purposes, on a pretext of ARS uses, is not allowed.[4] This exception is also subject to art. 89(1) safeguards, and 'subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject'. The controller must be able to demonstrate compliance (art. 5(2)).

(3) Art. 89(2)-(4) allows EU member state laws to derogate from data subject rights in relation to six articles (arts. 15, 16, 18, 19, 20 and 21), but only if and in so far as exercise of those rights would 'render impossible or seriously impair' the purpose of the processing for ARS uses. Therefore, EU law allows user rights to continue in relation to this processing, except in very specific circumstances. Providing some of these rights to pseudonymised data being held for ARS purposes may be expensive, but that is not in itself a proper basis for derogating from a right.[5]

(4) Recital 33 envisages that broad consent can be given for future scientific research purposes.

Some of the GDPR provisions which can affect ARS uses of personal data (unless there are derogations) include:

(i) The right of access (art. 15), and right to rectification (art. 16) are regarded as essential components of data protection (and mentioned in the EU Charter of

[3] For example, EDPS Preliminary Opinion', p. 23

[4] EDPS Preliminary Opinion', pp. 23-24

[5] EDPS Preliminary Opinion', p. 21.

Fundamental Rights), so derogations will only be allowed in very exceptions circumstances

(ii) The right to restriction or suspension of processing (Article 18) will apply, subject to any derogations.

(iii) The right to erasure (including to be forgotten) (art. 17) will apply, without any derogation applying.

(iv) The right to object (art. 21(6)) has a special proviso that that, where there is processing for ARS purposes, 'the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her, unless the processing is necessary for the performance of a task carried out for reasons of public interest.'

(v) The limitations on automated processing (art. 22) will apply, without any provisions for derogation. In any event, processing for ARS uses is not consistent with use of data to make decisions about individuals.

1.4. Is commercial 'scientific research' by private companies permitted as ARS uses?

The processing must first qualify as 'scientific research'. The test of what is scientific research should require that research is the purpose of the processing. It is not enough that the processing uses 'scientific methods'. As the EDPS puts it 'For a controller to simply claim to process data for the purposes of scientific research is not sufficient'[6]. EDPS follows the former Article 29 Working Party which, in its guidelines on consent, understood scientific research as a 'research project set up in accordance with relevant sector-related methodological and ethical standards', which would normally require 'informed consent, accountability and oversight'. Under this approach, only scientific research performed within an established ethical framework would therefore qualify as activities falling within the GDPR's special data protection regime for scientific research.

GDPR Recital 159 refers to 'privately funded research' as part of 'scientific research',[7] while saying this should be 'interpreted in a broad manner'. It does not explicitly refer to 'private sector invested research' or research by the private sector. However, the EDPS concludes that 'profit-seeking commercial companies can carry out scientific research' for GDPR purposes.[8]

In both the GDPR articles and recitals, the meaning of 'scientific research' is not very clear, but it would probably be limited to processing which is primarily for the purpose of research, and not primarily for some other purpose (such as routine commercial processes) and only incidentally for research. As the EDPS concludes, 'the research is carried out with the aim of growing society's collective knowledge and wellbeing, as opposed to serving primarily one or several private interests'.[9]

1.5. Safeguards are necessary for ARS uses, based on 'data minimisation'

'Data minimisation' is the key element of safeguards. Recitals 156-163 give details of the intentions of the GDPR in relation to processing for ARS uses. 'Member States should [provide]

[6] EDPS Preliminary Opinion', p.12 seems uncertain whether this is so.

[7] Recital 159: 'the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research'.

[8] European Data Protection Supervisor *A Preliminary Opinion on data protection and scientific research*, 6 January 2020, p. 11 (hereinafter 'EDPS Preliminary Opinion').

[9] EDPS Preliminary Opinion', p.12.

specifications and derogations with regard to the information requirements and rights to rectification, to erasure, to be forgotten, to restriction of processing, to data portability, and to object … [including] specific procedures for data subjects to exercise those rights if this is appropriate … along with technical and organisational measures aimed at minimising the processing of personal data in pursuance of the proportionality and necessity principles' (Recital 156).

Pseudonymisation is mentioned, but is not necessarily sufficient. Art. 89(1)(1), which provides that the safeguards provided in processing for ARS uses 'may include pseudonymisation' if that will allow the purposes of the processing to be fulfilled, but it also provides that 'Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes **shall** be fulfilled in that manner.' In other words, if **complete anonymisation** can fulfil the purposes of processing, it must be used, not just pseudonymisation.

1.6. How is disclosure to third parties, or publication, justified as ARS uses?

The GDPR's definition of 'processing' includes disclosures to third parties, but this is not dispositive. Processing of personal data for ARS uses must meet other criteria which disclosure to third parties might not satisfy. **Disclosure** of pseudonymised data to third parties increases the risk of misuse / data breaches, and therefore fails to provide sufficient safeguards. For whose purpose (which must be an ARS purpose) is the disclosure occurring? It must be the purpose of the data controller, not the third party. The receiving third party would need its own new lawful basis for the processing, under art. 6(1) as initial processing, not as further processing under art. 6(4).

Publication of pseudonymised data is too dangerous, because the subsequent uses are uncontrollable, and the risks of subsequent re-identification are not preventable. Publication of pseudonymous data cannot ever satisfy art. 89.

1.7. Can databases be linked by pseudonyms for ARS uses?

The GDPR does not address specifically the use of pseudonyms to link databases in relation to processing for ARS uses. Recital 157 refers to 'coupling information from registries', which implies that this is a reference to linking data in relation to individuals across multiple registries. Recital 157 clearly supports such linking, but does not state that it is presumed to be lawful in all cases. How can an art. 5(2) presumption of compatibility operate when the different databases are collected for different purposes? The Art. 6(4) tests of compatibility seem appropriate, or alternatively a new art. 6(1) ground of lawful processing might be needed. In both cases, recital 157 would provide support, but compatibility would not be assumed.

Such linkages would probably constitute profiling, even if the data in each registry is pseudonymous, and so all GDPR provisions relating to profiling would apply, including art. 21 on the right to object, art. 22 on automated processing, and the guidance of the EDPB on profiling (Recital 72).. ARS uses are, however, unlikely to result in breaches of the GDPR.

1.8. Can sensitive data be included in ARS uses?

Art. 9(2)(j) allows processing of any type of sensitive data for ARS uses, but only under conditions:

(i) **'necessary'** for ARS purposes
(ii) art. 89(1) observed – safeguards (data minimisation)
(iii) 'shall be **proportionate** to the aim pursued';
(iv) shall '**respect the essence** of the right to data protection'; and

(v) shall 'provide for suitable and specific **measures to safeguard** the fundamental rights and the interests of the data subject'.

The last three requirements must be implemented in Member State laws (by analogy, Korean laws in this context). The EDPS also concludes that the GDPR 'requires adoption of EU or Member State law before the use of special categories of data for research purposes can become fully operational'.[10] The ability to use sensitive data for research does not arise simply from the terms of the GDPR, it requires implementing EU or Member State legislation. The inclusion of sensitive data in ARS uses, should therefore not be assumed, but explicitly provided for in legislation.

Use of sensitive data for research therefore either requires separate procedures for sensitive and non-sensitive data, or non-sensitive data must meet the same standards. The presumed compatibility of ARS processing of sensitive data already collected (art. 5(1)(b)) applies to sensitive data, so that there is no need to satisfy the art. 6(4) compatibility test.

The rest of the GDPR also applies to such uses of sensitive data, as discussed previously in relation to other personal data. In addition, art, 22(4) places tight limits on automated processing of sensitive data, but this should never be relevant to ARS uses.

1.9. Penalties for breach – higher administrative fines; compensation

Infringement of the requirements of art. 89 or 9(2)(j) may result in the higher level of administrative fines (up to 20M euros or 4% of global turnover in the previous financial year, whichever is higher). Breaches may also result in exposure to possible compensation actions by data subjects.

Conclusions: Korea in EU shoes

The GDPR does allow some secondary processing of personal data held by a controller without need to satisfy the 'compatible uses' test of art. 6(4) (the 'presumption of compatibility'), but only under various sets of conditions, and the continuing operation of other provisions of the GDPR:

1. Pseudonymous data is still personal data, and not *per se* exempt from art. 6(4);
2. Only four categories of uses of personal data are so exempt: 'archiving purposes in the public interest, scientific or historical research purposes or statistical purposes '('ARS uses'), with no implied extension to similar uses (no 'etc');
3. The GDPR's provisions still apply to ARS uses of pseudonymised data, with only four exceptions: no need for a separate lawful basis for processing; longer storage where solely for ARS purposes; and derogations from six rights, only if provided by Member State laws, and to the extent essential for ARS uses; broad consent can be given for future scientific research purposes..
4. Any uses of pseudonymised data which are not ARS uses, including use to make decisions affecting individuals, must satisfy the art. 6(4) 'compatibility test'.
5. Commercially-funded scientific research may be allowed, but is limited to processing which is primarily for the purpose of research, not for primarily commercial purposes. Use of 'scientific methods' is inadequate, if the research purpose is missing.
6. Safeguards based on data minimisation are required for ARS uses; they may go beyond pseudonymisation, and require anonymisation, encryption etc.

[10] EDPS Preliminary Opinion, p. 17.

7. Disclosure of pseudonymised data to a third party increases the risk of lack of safeguards, and the third party must have their own lawful purpose for using the data; Publication of pseudonymised data cannot comply.
8. Linking multiple databases for ARS uses using pseudonyms is supported (Recital 157), but compatibility might not be presumed (uncertain). Compatibility may need demonstration (art. 6(4), or new lawful processing shown (art. 6(1)).
9. Processing of any type of sensitive data for ARS uses is allowed (in theory), but only under conditions which must be met for each example of use, and may be different from the conditions for use of non-sensitive data;
10. Failure to meet these conditions exposes a controller to the higher level of fines, and possible compensation actions by data subjects.

Overall conclusion: pseudonymisation of personal data for the purpose of ARS uses does not provide a blank cheque for use of the data.