The new General Data Protection Regulation -
Is there sufficient pay-off for taking the trouble to anonymize or pseudonymize data ?

Waltraut Kotschy,

Ludwig Boltzmann Institute for Human Rights, Vienna

A. DEFINING ANONYMIZATION AND PSEUDONYMIZATION IN THE PRESENT AND IN THE FUTURE EU LEGAL FRAMEWORK FOR DATA PROTECTION

When dealing with the topic of this article it seems necessary to start out by exploring if and how the GDPR defines "anonymization" and "pseudonymization" : Looking through the text of the GDPR we find a definition of "personal data" in Art. 4 (1), further a definition of "pseudonymization" in Art. 4 (5), but no definition of "anonymization" although very important consequences are attributed to anonymization: "Anonymized data" are outside the remit of data protection. An explicit reference to this effect in contained in recital 26 of the GDPR, stating that, "the principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person". "Anonymized data" are – according to the same recital – "personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."

Compared to the present legal situation under the Directive 95/46 the concept of "pseudonymized data" is new. By including a definition in the GDPR the longstanding discussion is ended, whether "pseudonymized data" are "personal data": they certainly are[1], and therefore are subject to the GDPR, which constitutes one of the essential differences in comparison to "anonymized data".

The definition of "pseudonymization"[2] of personal data, given in Art. 4 (5) GDPR, shows that the result of pseudonymization is the conversion of data about an identified person into data about a merely "identifiable" person. Condition is, that the additional data necessary for re-identification, are kept safely inaccessible for the users of "pseudonymized data".

As to the meaning of "identifiable" we have to look into the concept of "personal data". Concerning this key concept of data protection the GDPR has not brought about significant change; the definition of "personal data" in the GDPR is nearly identical to the definition contained in Directive 95/46 [3]: Still everything depends on what is "identifiable", and this finally depends on what is "identified", as someone is identifiable if he or she "can be identified, directly or indirectly, ….". This definition was

---

[1] Rec. (26) "…….Data which has undergone pseudonymisation, which could be attributed to a natural person by the use of additional information, should be considered as information on an identifiable natural person. ….."

[2] "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;"

[3] Art 4 (1) GDPR: "personal data' means 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* or to one or more factors specific to the physical, physiological, *genetic*, mental, economic, cultural or social identity of that natural person;" (alterations italicized by the author)
Art 2 (a) Dir 95/46: "(a) 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is 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;".

and is not terribly helpful, as it does not really say, when somebody has to be considered as "being identified". The Art. 29 Group tried to establish more clarity concerning the concept of "personal data" in its Opinion 4/2007[4]: Two of the most important results of this paper are[5]

- a meaningful definition of "identification" as the description of a person in a way which "singles this person out" from all other persons under scrutiny, and
- the insight, that "identification" and "identifiability" are not something absolute, but depend "on the circumstances of the case".

Even if it is necessary for practical reasons to develop an understanding of "(non-)personal data" which is generally applicable and not entirely a "case to case"-affair, the contextual approach, proposed by the Art. 29 paper, shows the possibility to take account of certain categories of use-situations, where the risk of re-identification is lower than in others: In case of the publication of anonymized or pseudonymized data the risk of re-identification is always higher than in case of the use of such data in an environment where the users as well as the circumstances of use are limited, so that actual risks can be better assessed and counteracted. Even if the risk of re-identification of the data subject cannot be entirely abolished by pseudonymization, or even anonymization, as we have been told time and again[6], we can at least reduce this risk significantly by choosing only reliable recipients of such data and by preventing misuse – in form of re-identification - by adequate safeguards.

B. AN EXISTING EXAMPLE FOR PRIVILEGED USE OF PERSONAL DATA WITH DISGUISED IDENTITIES

The idea of distinguishing between different risk levels when allowing for the further use of data has been specifically taken up in the context of implementing the Data protection Directive 95/46 in Austrian.

First of all, the Austrian Data Protection Act (DSG 2000) introduced pseudonymized data under the denomination of "indirectly personal data". This wording acknowledges that

- these data are still "personal data" and that
- they identify a person, however only "indirectly", in the sense, that additional information would be needed to reveal the full identity of the data subject.

This concept had been developed in cooperation with the research community in Austria, especially representatives of the medical universities. They explained that anonymized data usually are not sufficient, partly because the informative value has to be reduced considerably in the course of anonymization, partly – and even more importantly – because checking on accuracy of data and information on further development of "a case" make it necessary to use disguised identities which can be linked, in case of need, to the "true" identities by the initial source of the data. The need to check on the accuracy of data turned out also to be of prime interest for the Austrian official Statistics Bureau, which were very keen to assert, that they could never work efficiently with anonymized data.

The Austrian Data Protection Act 2000 contains specific regulation on the nature and the use of "indirectly personal data":

> 1) Such data relate visibly to one specific person, however, all identifiers, which *together* directly identify this person (such as the name plus date of birth plus residence etc., are

[4] Opinion 4/2007 on the concept of personal data, WP 136, from June 20th 2007
[5] P 13 of WP 136
[6] As a prominent example see the Art. 29 Group's Opinion 05/2014 on Anonymisation Techniques, WP 216 from 10 April 2014

encrypted and the user of such data has no access to the encryption algorithm.[7] Consequently the algorithm must be "sufficiently safe" according to the state of the art.

2) "Indirectly personal data" may be transmitted to and used by third parties without the restrictions of Art. 7 of Directive 95/46: As far as they are used for research and statistics purposes they represent the implementation of Art. 6 (1) (b) of the Directive, dealing with "further use for not incompatible purposes" under adequate safeguards foreseen in national law.

3) Using "indirectly personal data" triggers – under special conditions as listed below - several privileges for the controllers involved: *inter alia*

- no obligation to notify the processing of indirectly personal data to the DPA,
- no restriction for disclosing such data to third parties,
- no obligation to obtain permission from the DPA for transfers to third countries,
- no obligation to inform the data subjects about transfers to third parties,
- access rights of data subjects are suspended

Special conditions respectively provisions for granting these privileges are:

- the recipient assures that he will refrain from any attempts to re-identify these data
- the recipient is known to be reliable
- activities to the contrary on his part are specifically punishable.

These rules have been in force since 2000 and have proven practicable, especially so in the area of providing data for scientific research and statistics. Cases of misuse have not become known so far. Official statistics in Austria have developed a system of doing census by means of "indirectly personal data", which has several enormous advantages:

- instead of having to ask the citizens about the relevant data, the necessary data are taken from existing registers of public administration. Linking the relevant data about one data subject is done in a data protection friendly way: they are not linked by means of name, date of birth etc. but by means of a specially created pseudonym;
- instead of doing census only every 10 years, statistical surveys using data from registers which are part of the system (- they are defined by law-) can be made whenever necessary.

C. REFERENCE TO THE EFFECTS OF ANONYMIZATION AND PSEUDONYMIZATION IN THE GDPR

1. As **anonymous data** are not the subject of data protection, the text of the GDPR does not openly refer to them; only in some recitals such data may be mentioned pointing out differences to the significance of personal data.[8] All we gather from the GDPR about "anonymized data" is that they do not relate to an identified or identifiable person and are therefore irrelevant for matters of data protection.

2. Where are consequences of "pseudonymization" mentioned in the GDPR?

a) - **Art 89 (1):** mentions pseudonymization as a means to enhance protection in the course of processing data for archiving purposes in the public interest, or scientific and historical research purposes or statistical purposes.

[7] This is why the initial controller, who has rendered personal data to be "only indirectly personal", profits from the privileges for processing indirectly personal data only in so far as he may disclose such data to third parties under condition of their reliability concerning non-identification

[8] See e.g. rec. 26

- **Art. 6 (4) (e)**: use of data in pseudonymized form *may* contribute to the compatibility of further use (and thus allow for privileged further use without need for special consent or legal provision).

- **Art. 9 (1) (j)**: processing of special category data for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is lawful only "in accordance with Art. 89 (1) based on Union or Member State law…"

  By mentioning Art. 89 (1) this provision relates also to pseudonymization, which is obligatory if the purpose allows for using pseudonymized data.

- **Art. 25**: Data protection by design, e.g. by means of pseudonymization. Art. 25 contains a very generally formulated obligation for the controller to establish "data protection by design". The implementation of such design by means of using pseudonymized data could reduce the risks of processing which could result in a positive outcome of the necessary data protection impact assessment (art. 35).

In all these examples pseudonymization is mentioned as a means for making data processing legal in cases which would otherwise not be lawfully possible. HOWEVER, the text of the GDPR is rather ambiguous concerning the question of whether pseudonymization *alone* achieves lawful processing: The texts suggest that it is rather just a contribution to a mix of criteria which is necessary as a whole to make processing lawful:

> "The application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations. The explicit introduction of 'pseudonymisation' in this Regulation is not intended to preclude any other measures of data protection."[9]

No distinction is made in the provisions of the GDPR between

- a controller who uses "pseudonymized data" for some processing operations, but is in possession of the – separately kept - re-identification mechanism, and
- someone, especially a third party, who has no access to this mechanism (-using one-way cryptography should be included in this case scenario):

In terms of possible misuse the latter situation is clearly less "unsafe" than the first one. None of the provisions of the GDPR reflects on these differences in a clearly recognizable way. On the contrary: Rec. 29 explicitly states that "pseudonymization" should also be recognized as a means for enhancing data protection if it is applied "within the same controller", which evidently means: who is in possession of the decryption mechanism. Whereas it has to be acknowledged that pseudonymisation also has positive data protection effects when used "within the same controller", clear privileges can only be attached to situations where there is practically no risk. Therefore:

- ➔ The potential of pseudonymization in data protection could be enhanced if especially transfer to and use by "third parties", that is controllers who have no access to the decryption algorithm, were clearly privileged.
- ➔ A commitment by the recipient not to counteract pseudonymization, together with severe punishment if such commitment is violated, should be included into the conditions for privileged use of pseudonymized data.

[9] Recital 28 to the GDPR

One of the big questions at present is, whether there is still enough room of manoeuvre within the national application of the GDPR to establish more precise rules concerning the effects of pseudonymisation. Art. 89 (1) GDPR does not show any reference to national legislation for specification. Whether processing of sensitive data for research or statistical purposes should or must be regulated in more detail even than is done by Art. 89 (1), is not entirely clear: Art. 9 (1) (j) speaks of processing "in accordance with Art. 89 (1) based on Union or Member State law". This seems to demand that two kinds of conditions must be fulfilled e.g. for medical research: Data must be pseudonymized whenever possible AND there must be a special provision in Union or Member State law which allows for the use of health data for the purpose of research.

This latter condition would most likely be fulfilled e.g. in Austria by the Law on Medical Universities which does not only allow for the use of health data for research purposes but positively obliges the medical staff of medical universities to do so. For other areas of research such legal provision might, however, be missing, which may be the case also in other Member States, considering the fact that there is also freedom of scientific research.

At present, according to the Austrian legal framework, it is sufficient that "only indirectly personal data" are processed, that is: safely pseudonymized (medical) data, which are not likely to pose a risk for the data subjects. The future legal situation may be a real hindrance for research activities in some cases as safe pseudonymisation alone will not be sufficient. It seems deplorable that here again pseudonymisation will not create a clear advantage: Whether doing research by means of pseudonymized health data will be lawful, is up to further regulation, be it by the Union or – even worse – by each Member State.

b) - **Art. 11** and **Art.s 15 – 20**: In case of processing data without identification of the data subjects, the controller is not obliged to store or keep information which would enable re-identification just for the sake of being able to answer to requests from individuals according to Art. 15 – 20 (rights of the data subjects). However, if the data subject, for the purpose of exercising his or her rights under those articles, provides additional information enabling his or her identification, the controller has to respond faithfully to the requests of the data subject.

.

Processing of pseudonymized data would thus result in reducing the obligations for controllers under Art. 15 – 20 of the GDPR in many cases.

Austrian experience since 2000 with general suspension of access rights to "indirectly personal data" has been showing no problems in this respect.

D. CONCLUSIONS:

Using anonymized data results in clear consequences under the GDPR: The GDPR is not applicable. So, using anonymized data will "pay off" under the Regulation, but there is always a risk that "anonymization" has not been fully achieved. Although the consequences are clear, the requirments for dealing with truly anonymized data are less clear.

Using pseudonymised data under the GDPR

- ➔ has no precise legal consequences: Only on a case to case basis it can be evaluated whether a processing operation is rendered lawful by means of using pseudonymized data;

- ➔ Using pseudonymized data does not induce clear and immediate legal advantages, such as e.g. privileged transfer to third parties and/or to third countries.
- ➔ The potential "pay-off" for pseudonymization in data protection has not (yet) been fully exploited; it will have to be seen, whether the instruments foreseen in the GDPR for establishing more detailed rules, such as for instance Codes of Conduct, could be made use of: If in such instruments the conditions could be defined and listed which would free the processing of pseudonymized data from further restrictions, the trouble of pseudonymization would truly "pay off" for the users and also for the data subjects as the risk of processing non-pseudonymized personal data would be avoided to a higher degree.