Identifiability: Policy and practical solutions for anonymization and pseudonymization

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

Waltraut Kotschy, Vienna

A. CLARIFYING CONCEPTS

What is anonymization vs. pseudonymization?

Anonymised data:

A set of data which is meant to *relate to one single individual*[1], whose identity is, however, not recognizable: Any possibly identifying features have been removed from this set of data. As a consequence nobody should be able to recognize the actual data subject.

PROBLEM 1: How far can features be removed and the information still kept meaningful?

- ➔ It is often upheld, that data are scarcely ever safely anonymized, as it might be possible to find additional information which evidently belongs to the same data subject and could even lead to identification of this data subject

PROBLEM 2: Whose knowledge or ability to find additional information shall be the benchmark for "safe" anonymization? Is it enough to refer to a user with average knowledge and abilities?

Pseudonymized data:

If "adding on" to data about a certain individual or "re-visiting" such data (e.g. for purposes of checking plausibility) will foreseeably be necessary, pseudonymized data will have to be used instead of anonymized data. Pseudonymized data are data which recognizably relate to a specific data subject, whose identity has, however, been disguised (encrypted). If the decryption algorithm has been lost or destroyed, pseudonymized data become functionally anonymized.

"Safety" of pseudonymization depends on two items:

- Firstly, on whether additional information could be found which makes identification possible → the same problem as with anonymization
- Secondly, on the quality of the encryption of the identifiers.

Likely, there is no absolutely safe way of encrypting identities

- ➔ Besides encryption of the identifiers, are there complementary measures which could create an overall situation of using pseudonymized data which is "sufficiently safe"?

[1] In contrast „aggregated data", which are also often referred to in the context of anonymization, are not meant to refer to one individual, but to a whole number of individuals.

B. AN EXISTING EXAMPLE FOR PRIVILEGED USE OF PSEUDONYMIZED DATA

Austrian Data Protection Act 2000 contains specific regulation on the use of pseudonymized data:

1) In order to qualify in this context, the chosen method of pseudonymization must be "sufficiently safe" according to the state of the art (→ "indirectly personal data")

2) using "indirectly personal data" triggers several privileges for the controller: *inter alia*

- no obligation to notify processing to the DPA
- no restriction for disclosing 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 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
- actions to the contrary on his part are specifically punishable.

3) 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.

C. REFERENCE TO ANONYMIZATION AND PSEUDONYMIZATION IN THE GDPR:

1. Definitions:

The GDPR defines only "pseudonymization": Art. 4 No 5.

> "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 about "anonymized data"? They also "can no longer be attributed to a specific data subject." According to the GDPR the consequences of these different categorisations are decisively different:

- ➔ "pseudonymized data" are still "personal data" and therefore subject to the rules of the GDPR, whereas
- ➔ "anonymized data" are outside the remit of the GDPR.

> 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. …..
>
> Rec. (26) ……..The principles of data protection should therefore not apply to anonymous information, that is information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is not or no longer identifiable. This Regulation does therefore not concern the processing of such anonymous information, including for statistical and research purposes.

The fact that personal data, which have been rendered anonymous, are no longer covered by data protection rules, seems justified only for *safely* anonymized data.

➔ Especially under the GDPR we need standards clearifying what is "(sufficiently) safe" anonymization and pseudonymization.

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.

- **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. 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 these cases pseudonymization is named 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 (in all cases) 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 (see also Rec. 28)

No distinction is made in these provisions whether pseudonymized data are used

- by a the controller who uses "pseudonymized data", but is in possession of the re-identification mechanism, or
- by a third party which does not possess 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: see Rec. 29)

➔ 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.

There is most likely still enough room of manoeuvre within the national application of the GDPR to establish such rules.

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.

CONCLUSION: Using pseudonymized 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 third countries.
- ➔ The potential "pay-off" for pseudonymization in data protection has not (yet) been fully exploited

D. THE FUTURE E-PRIVACY RULES

One of the main reasons for having new European DP rules was better response to the special data protection issues connected with the use of the internet.

However, → the most important special data protection rules for internet communications are contained in the e-privacy Directive.

EU-Commission announced that a new e-privacy Directive shall be created – relationship to the GDPR?

At present: relationship between GDPR and e-privacy Directive described in Art. 95 of the GDPR:

> "Article 95
>
> **Relationship with Directive 2002/58/EC**
>
> This Regulation shall not impose additional obligations on natural or legal persons in relation to processing in connection with the provision of publicly available electronic communications services in public communication networks in the Union in relation to matters for which they are subject to specific obligations with the same objective set out in Directive 2002/58/EC."

However, what about lessening the obligations of controllers in cases where the GDPR would allow for it, as e.g. in case of use of pseudonymized data?

What is e.g. the relationship between Art. 6 (4) GDPR (on compatible further use) and Art. 5 (3) of the present e-privacy Directive which requests consent for all forms of processing involving access to information in the terminal equipment of the subscriber.

Are data about subscribers who are identified only via their communication equipment "pseudonymized data" or not?

Could targeted advertising on the basis of tracking ever be "compatible further use"? If rather not, could this, at least, be the case, if the recipient of the ad-message is a former customer?

These are questions which are most relevant in daily commercial life and which should be discussed extensively in the context of readjusting the e-privacy rules to modern times.