PSEUDONYMOUS IDENTIFIABILITY AS A SOCIETAL PROBLEM

Ahti Saarenpää

Professor Dr. emeritus, University of Lapland, Institute for Law and Informatics Box 122
96101 Rovaniemi, Finland
ahti.saarenpaa@ulapland.fi

Keywords: ***Personal data, Pseudonymous data, GDPR, Openness, Public data, Fundamental rights, Self-etermination, Access, Flexible document***

Abstract: *The use of pseudonymous identifiers for various purposes is a longstanding practice, and attention was drawn to their legal implications quite some time ago. More often than not the interest lay in criminal law. Typically, the question asked was when using a false name was permitted or prohibited. It was particularly important for an individual to use his or her correct identity when interacting with public authorities.*

The European General Data Protection Regulation (GDPR) has cast the use of pseudonymous data in a new light legally. It contains provisions dealing explicitly with the use and processing of pseudonymous data.

Legislation to date has left us with the impression that anonymous data is the opposite of personal data. The concept of personal data is a broad one: it is data that enables identification directly or indirectly. The crucial consideration here is that an individual can be identified. This must not be circumvented through pseudonyms. Anonymous data, however, is data that makes it impossible to identify individuals. To this day data protection legislation has never applied to the processing of such data. This makes the concept of anonymous data a very important one legally.

The GDPR made pseudonymous data an explicit aspect of European data protection legislation. With the Regulation now in force, the use of pseudonyms is an integral and default procedure in implementing data protection as well as information security. Using pseudonymous data enables us to reduce the risks relating to the processing of personal data, but may cause a range of problems where openness is concerned in our modern network society. What is good for data protection is not always so good for other constitutional rights, especially openness.

My article reflects on the benefits and liabilities involved in the use of pseudonyms. With a view to the functioning of the public sector, it delves shortly into the impacts that the principle of public access to official documents might have on the use and processing of pseudonymized data.

1. From a bipartite to tripartite classification of personal data

Legislation on personal data is legislation that regulates identifiers and their processing. Personal data is an identifier. This sums up the concept of personal data we have worked with and the concept is quite a broad one. Data that makes it possible to identify an individual indirectly also generally falls under the definition of personal data found in legislation.

This consideration is often forgotten where efforts are made to circumvent the application of personal data legislation to the processing of data.[1] One need read no farther than *the recitals* of the Personal Data Directive to see the emphasis on how broad the concept of indirect data is. Personal data is more than specific identifiers. The same consideration appears in recital 26 of the General Data Protection Regulation (GDPR or only Regulation), where the point of departure is to take into account the use of *all objective means* to identify an individual. The dichotomous opposite of personal data is *anonymous* data. The latter refers to data which provides no possibility of identifying an individual using objective means. This being the case, such data fall outside the scope of data protection legislation. Genuinely anonymous data or data that have been rendered such through processing are not personal data.

Yet this essentially straightforward distinction occasionally causes confusion. One simply must remember that in an information system anonymity must affect everything. The fact that a controller provides another party with data in anonymous form is not sufficient.[2] Such a procedure does not release the controller from its obligation to comply with data protection provisions. Even if the data are anonymous when in the possession of the recipient, the delivery of the data must meet the requirements set out in the law.

The third significant category of data in general terms and in the context of the Regulation is *pseudonymous* data. Its use was possible even under the Personal Data Directive. The Directive does not make explicit mention of such data but there was nothing that would have prevented its use at the time in processing personal data unless national legislation contained provisions to the contrary. Art. 29 Working Group, consisting of data protection officials, in fact examined pseudonymous data as one form of personal data in a 2004 opinion on personal data.[3]

In the Regulation pseudonymous data is an important type of data, one that is processed separately. It is mentioned in the recitals and articles a total of 15 times. The number of times a term occurs is rarely a particularly important consideration, but the present case is an exception. Its frequent occurrences in the recitals indicates the importance of pseudonymous data; moreover a definition, is provided in the operative part of Regulation.[4] It reads as follows:

> «‹pseudonymisation› means 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;»

This definition is in principle rather clear. Pseudonymous identifiers are used to avoid individuals being identified by third parties and in the operation of the controller by persons whose duties do not involve the processing of the data in question. If we take a closer look at the Regulation, we can also observe a desire on the part of the legislator to make pseudonymized data an important part of the Regulation's data protection system. I will now go on to examine this system-level question in more detail.

---

1 One of the more striking instances of incompetence in Finnish practice was a case where a regional government authority used unsecured email when asking a doctor for information on the health of a public official and included the official's personal identity code in the message as an identifier. The information was considered anonymous.

2 In 2017, the Finnish Data Protection Board considered an application in which a university hospital reported that it was sending patient information in anonymous form to a research institute in the United States. The hospital thought that this would obviate the need to apply data protection legislation in the case. However given that the hospital still had the original material, which included personal data, the anonymization cannot be considered genuine.

3 Opinion 4/2007 on the concept of personal data.

4 In light of the unfortunate wordiness style of the Regulation, the number of occurrences of a term is not a sufficient criterion for establishing its importance. An objective assessment of the term's importance would require the use of other criteria as well.

2. Pseudonymous data in an information system

The definition of pseudonymous data is prominently enhanced by making it part of *data protection by design and data protection by default.* The concept may be new on the legislative level but the basic idea is an old one.[5]

Article 25 of the Regulation mentions pseudonymous data as an example of organizational and technical solutions that serve the implementation of data protection. The inclusion of the concepts of data protection by design and data protection by default in the actual text of the Regulation is interesting and exceptional in terms of legislative technique. In traditional regulation, *examples* are typically presented in the rationale for statutes, not in individual provisions. Regulation departs from the practice. Article 25, as we can see relies squarely on examples:

> «Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.»

As the use of examples in legal provisions is not, to my knowledge, common in most countries, it was a comparatively courageous move to elevate pseudonymization to a significant data protection principle solely by dint of a definition and what is written in Article 25. However, if we look at the article in light of many of the statements in the recitals, the situation changes. The reader is left with the distinct conviction that the legislator has sought to make pseudonymization a strong principle in the implementation of data protection, one to be taken seriously. As Professor *Peter Blume* has aptly put it, Article 25 is *a wake-up call*.[6] Pseudonymization most definitely needs the precision provided by sound legislation, but the legislation can also convey a special message on the importance of regulation.[7]

But what kind of a wake-up call is Article 25 ultimately? And can we speak as we typically do of the literal interpretation of the text of a statute? The style in which the Regulation is written is quite challenging. The simple «Open the law book» approach will not do the reader much good here.[8] We have to know something more much more about the legal system of which pseudonymization is a part. Interpretation, as Professor *Aulis Aarnio* has most appropriately pointed out, must be in tune with the legal system at hand. I would add here that we should speak of the *intended system*.

Despite its reliance on examples, it is my impression that Article 25 is telling us that pseudonymization is not merely an option. It is more. Read in conjunction with recitals 28, 29, 78 and 125, I would say that the article indicates that what we see is a strong main rule. The *main rule* is to pseudonymize, the *exception* is to use open data.

Recital 18 states clearly that applying pseudonymization in the case of personal data can reduce the risks to the subjects involved and help the controller and processors of the personal data comply with their data protection

---

5 Finnish Data Protection Ombudsman Reijo Aarnio has repeatedly emphasized that the notions of data protection by design and default have been elements of good data file practice from the outset.

6 See Blume, Smart Data Protection, p. 180.

7 See also Bygrave, Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements, Yulex 2017, pp. 42–58.

8 The expression «Open the law book doctrine» is one used by Professor Kauko Wikström to refer to the methodological routine use of legal texts in our daily legal life.

obligations. In this context, pseudonymization is one of the available tools in the effective implementation of data protection. Similarly, Recital 29 speaks of incentives encouraging the use of pseudonymizations. Recital 78 is rather general in nature but it nevertheless contains an important message that controller should implement pseudonymization as soon as possible in fulfilling their data protection obligations. Recital 156 for its part makes pseudonymization a crucial safeguard to be used in the processing of personal data for archiving purposes in the public interest. Taken together, all of these statements, to my understanding, justify seeing pseudonymization as a main rule in the lawful processing of personal data. Further support for this view can be seen in Article 6 where pseudonymization is mentioned as a safeguard comparable to encryption. This highlights the renewed and crucial link between data protection and information security.

In my view, this interpretation prompts the conclusion that when planning, acquiring and using information systems a controller must undertake to determine and demonstrate whether it is at all acceptable to operate without pseudonymization or anonymization. The core issue here is the assessment of qualitative risks, a consideration that affects the entire path information takes, not merely the processing of a particular set of data by the controller in keeping with the purpose of the data file.[9]

In light of this assessment, pseudonymization now featuring prominently as an alternative to anonymization and processing open identifiers emerges as one of the cornerstones of the renewed European data protection system. It clearly links data protection and *information security*. This link was comparatively weak in the Data Protection Directive and granted quite a bit of discretion and thus liberties to controllers. Now, there is no wiggle room allowing us to read Article 25 without bearing in mind that *good data protection requires sound information security*. There are thus no two ways about it: we must take pseudonymization seriously. It is far more than what might seem a one-off example in Article 25. But we do not always seem to remember this.[10]

Pseudonymization is not a new practice. It has been possible to use it and it should have been used earlier as part of *good data processing* practice. In the literature I have cited a regrettable case in Finland involving a *leak of sensitive data*.[11] The National Institute for Health and Welfare inadvertently placed the health data of over 6000 people on an information network, the Internet. The leak was not discovered by the Institute; they found out about it by notice of the national Data Protection Ombudsman. A concerned citizen had alerted this office.

When the Institute ultimately made a public statement about the incident over a month later, the reason given was «human error». One of the people working there had used information containing personal identity codes when putting together a report. He or she then, without thinking about it, put the information on an open network.

I think there are three salient observations to be made here from the legal point of view. First, the data protection legislation prescribes that when processing personal data, every effort should be made to avoid processing unessential data. This means avoiding the use of identifying data that could be exploited by unauthorized parties. Today this means the use of pseudonyms.

Secondly, information systems should always be planned to minimize the risk of human error and the impacts of such error if it occurs. This clearly had not been done in the case. Some «dummy» managed to make a major mistake because there were no safeguards in place in the system that would prevent such errors or alert users to potential problems.

---

9 See also Tarhonen, Pseudonymisation of personal data according to the General Data Protection Regulation, pp. 25–26, in: Edilex, 15 November 2018.

10 For example, when discussing Article 25 in his textbook, Peter Blume unfortunately does not take up pseudonymization in any detail. See Blume, Sen nye persondataret, pp. 143–145.

11 See also Saarenpää, Legal Informatics and the Scarsity of Justice, in: Schweighofer / Kummer / Saarenpää / Schafer (eds.), Datenschutz LegalTech, IRIS2018 proceedings, pp.397–398.

Thirdly, I would point out that information security is a core value in the Network Society and in the GDPR. It is one of the most essential forms of security in a society. The Institute for Health and Welfare had in no way sought to ensure that sensitive information would not end up on an open information network via an individual employee's computer. This was an instance of the organization neglecting information security.[12]

I will not go any deeper here and analyse the structural significance of pseudonymization as a part of an information system. I think this has become apparent in what I have had to say thus far. One thing that merits a closer look at his point is pseudonymization from the perspectives of the principle of public access to official and the use of public-sector information

3. The relation between personal data public and public access in the age of the GDPR

Finland and Sweden joined the EU when drafting of the Personal Data Directive was in its final stages. This prompted an addition to the Directive, recital 72, allowing the principle of public access to official documents to be taken into account when implementing the Directive. Thus, the Nordic principle of *public access to official documents* was linked to the protection of personal data in a special way, one that has caused difficulties in legal interpretation. Indeed, in the public sector one has occasionally seen views maintaining hereby that the principle of public access would more important than the protection of personal data.[13] Yet in Finland, as elsewhere, the principle of public access has only superseded and only supersedes the protection of personal data when there are express provisions allowing it.

It is equally essential to recall the idea, familiar from the earliest days of data protection legislation, that public access to particular data in no way «exempts» subsequent processing of that data from the provisions of data protection legislation. Data is not dichotomous by nature. The processing of personal data that becomes public in keeping with legislation on public access is still subject to data protection legislation unless the legislation on public access expressly indicates otherwise.

This crucial consideration has been visibly driven home in a judgement of the European Court of Human Rights ruled on 27 June 2017. In the case, the Court took opinioned that a company that published publicly available information on taxation for commercial purposes, an activity not journalistic in nature, had violated individuals» privacy contrary to the Convention on Human Rights.[14]

When the new GDPR was adopted, one aim was to clarify the relation between personal data and the principle of public access to official documents. This was done by including a specific article to this end. The issue was too important to capture adequately in recitals only. The outcome was Article 86, which reads as follows:

> «Personal data in official documents held by a public authority or a public body or a private body for the performance of a task carried out in the public interest may be disclosed by the authority or body in accordance with Union or Member State law to which the public authority or body is subject in order to reconcile public access to official documents with the right to the protection of personal data pursuant to this Regulation.»

The article, whose purpose is described in general terms in recital 154, clearly marks a step forward from Recital 72 of the Directive. Recital 154 also elucidates the relation between data protection and openness:

> This Regulation allows the principle of public access to official documents to be taken into account when applying this Regulation. Public access to official documents may be considered

[12] See Saarenpää, Legal Informatics: a Modern Social Science and a Crucial One in 50 Yers of Law and IT, p. 26.

[13] Quite typically, when public sector actors describe the legislation governing their activities on their websites, they start with the Act on the Openness of Government Activities.

[14] Application no. 931/13, Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland; see also EJC C-73/07, Satamedia.

> to be in the public interest. Personal data in documents held by a public authority or a public body should be able to be publicly disclosed by that authority or body if the disclosure is provided for by Union or Member State law to which the public authority or public body is subject. Such laws should reconcile public access to official documents and the reuse of public sector information with the right to the protection of personal data and may therefore provide for the necessary reconciliation with the right to the protection of personal data pursuant to this Regulation. The reference to public authorities and bodies should in that context include all authorities or other bodies covered by Member State law on public access to documents. Directive 2003/98/EC of the European Parliament and of the Council leaves intact and in no way affects the level of protection of natural persons with regard to the processing of personal data under the provisions of Union and Member State law, and in particular does not alter the obligations and rights set out in this Regulation. In particular, that Directive should not apply to documents to which access is excluded or restricted by virtue of the access regimes on the grounds of protection of personal data, and parts of documents accessible by virtue of those regimes which contain personal data the re-use of which has been provided for by law as being incompatible with the law concerning the protection of natural persons with regard to the processing of personal data.

However, as we can see, the GDPR still leaves Member States quite a bit of discretion when enacting legislation on and interpreting the relation between protection of personal data and the principle of public access to official documents. Where the protection of personal data and public access meet in the EU, the outcome may thus differ a lot in different states. Unfortunately.

As the GDPR also applies to the public sector, it is reasonable to assume that in the long term it will result in more uniform data processing procedures in the public sector. This point has been emphasized in Nordic Countries especially by Dr. Tuomas Pöysti. At the end of the day, the issue here is trust in the actions of the public authorities in the Network Society.[15] It is natural to embrace this concept if the protection of personal data is considered important in the government apparatus and related legislative drafting. This was not the case when the Directive was earlier implemented. Quite the contrary: a significant number of derogations were enacted. Compliance with the GDPR has not gone all that smoothly either. One case in point occurred at the end of 2018, when it turned out that the Finnish Transport Safety Agency (Trafi) had opened up an online service allowing users to find out who had a valid driver's license.[16] However, the service provided more information than was necessary and this could be combined with other available information. This was a violation of the principle of minimization of data in Article 5 of the GDPR. One aspect of the incompetence involved could be seen in the senior management at Trafi initially telling the media simply that that the information in question was public.

Particularly odd and regrettable in this case was that only a short time earlier the key government expert bodies on information management *Vahti* (the Government Information Security Management Board) and *Juhta* (The Advisory Committee on Information Management in Public Administration) had given Trafi an award for its active efforts in promoting digital data protection and information security. This will also no doubt tarnish the credibility of the two bodies down a notch.

4. Pseudonymous data in the public sector

In my view, acknowledging the importance of pseudonymous data in data processing is a straightforward aspect of implementing data processing in government. All in all, pseudonymous data have become a significant

[15] See Pöysti, Trust in Digital Administration and Platforms, pp 147, in: 50 Years of Law and IT.

[16] At the beginning of 2019 Trafi became a part of the new Finnish Transport and Communications Agency (Traficom).

factor in in the processing of data in government. Such data can serve to avert any number of deliberate and inadvertent information leaks.

Where the principle of public access is concerned, one problem is how our right to public information can be implemented when information management uses pseudonymization. This is not a wholly novel concern. It is and must be an essential point of departure in any form of information management that we have the right to public information in its proper structural context. This is a matter of access to information systems in the appropriate manner. The systems must be designed and implemented such that the protection of personal data and public access required by the relevant legislation are both realized in the appropriate manner.[17] Here the GDPR plays a crucial role. It is a guide to creating the proper, appropriately protected path for selecting data for inclusion in stores of open data. This all requires the planning and implementation of *dynamic public documents*. Government that relies on information systems involves more, far more, than provising for traditional public access to documents.[18] In this connection pseudonymous data is an effective solution. But, of course, there should be standards showing, how to give citizens the possibility to get to know the legal borders of their practical access. Access principle should be based on sophisticated pillars.

5. Concluding remarks

We still see cases where the parties and public bodies involved have difficulties understanding how the protection of privacy and personal data interacts with the principle of public access to official documents. The protection of personal data is readily seen as primarily a technical concern and generally as a matter relating to information systems. Such a limited approach overlooks the development of our fundamental rights in the Network Society. To ask whether personal data protection is – or think that it is not a genuine fundamental right is utterly misguided. It is a fundamental right. We simply cannot ignore the basic idea enshrined in the European Charter of Fundamental Rights that privacy and the protection of personal data as two fundamental rights of equal standing.[19] We must remain mindful of this whenever discussing the use of pseudonymized data in different situations. The manner in which the GDPR, even with its shortcomings, highlights the importance of pseudonymized data as a type of data is certain to enhance the importance of personal data protection as a European fundamental right.

And once again I would like to point out the important relation between data protection and data security. No acceptable data protection solution is possible without a robust information security environment and information security solutions. Almost every day we witness the outcome of shortcoming in this area even large infrastructural gaps in administration in Europe, and in fact everywhere.

6. References

Blume, Peter, Den Nye Persondataret, Persondataforordningen, Databeskyttelseloven, 2 ed., Jurist og Økonomiförbundets Forlag (2018)

Blume, Peter, Smart Data Protection, in: Wahlgren (ed.) 50 Years of Law and IT, The Swedish Law and Informatics Research Institute 1968–2018, Scandinavian Studies in Law Volume 65, pp.175–190.

Bygrave, Lee, Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements, in: Bygrave Dobos (eds) Yulex 2017, Senter for rettsinformatikk, pp. 42–58.

[17] Saarenpää, E-justice and the Network Society, Some comments from the Finnish point of view in Cerbena, Cesar A. (ed), Brazilian and European perspectives on e- Justice Brazilian and European perspectives on e-Justice, Edition in Portuguese and English, Curitiba, Federal University of Paraná (2016) pp.216 ss.

[18] See also Magnusson Sjöberg, The Swedish Administrative Act and Digitalisation, pp.309–320, in: 50 Years of Law and IT.

[19] Cfr. Van der Sloot B., Legal Fundamentalism: Is Data Protection Really a Fundamental Right?, in: Leenes R. / van Brakel R. / Gutwirth S. / De Hert P. (eds), Data Protection and Privacy: (In)visibilities and Infrastructures, Law, Governance and Technology Series, vol 36. Springer (2017).

Magnusson Sjöberg, Cecilia, The Swedish Administrative Act and Digitalisation, in: Wahlgren (ed.), 50 Years of Law and IT, The Swedish Law and Informatics Research Institute 1968–2018, Scandinavian Studies in Law Volume 65, pp. 309–320.

Saarenpää, Ahti, E-justice and the Network Society, Some comments from the Finnish point of view in Cerbena, Cesar A. (ed) Brazilian and European perspectives on e-Justice Brazilian and European perspectives on e-Justice. Edition in Portuguese and English, Curitiba, Federal University of Paraná 2016, pp. 211–234.

Saarenpää, Ahti, Legal Informatics: a Modern Social Science and a Crucial One, in: Wahlgren (ed.), 50 Years of Law and IT, The Swedish Law and Informatics Research Institute 1968–2018, Scandinavian Studies in Law Volume 65, pp. 16–38.

Saarenpää, Ahti, Legal Informatics and the Scarsity of Justice, in: Schweighofer / Kummer / Saarenpää / Schafer (eds.), Datenschutz LegalTech, Proceedings of the 21st International Legal Informatics Symposium IRIS 2018, WEBLAW, Bern 2018, pp.397–398.

van der Sloot, Bart, Legal Fundamentalism: Is Data Protection Really a Fundamental Right?, in: Leenes R., van Brakel R. / Gutwirth S. / De Hert P. (eds.), Data Protection and Privacy: (In)visibilities and Infrastructures, Law, Governance and Technology Series, vol 36. Springer, 2017.

Tarhonen, Laura, Pseudonymisation of personal data according to the general data protection regulation, in: Edilex www.edilex.fi (15 November 2018).