What’s in a name: the conflicting views of pseudonymisation under eIDAS and the General Data Protection Regulation

Niko Tsakalakis, Sophie Stalla-Bourdillon and Kieron O’Hara

Abstract:
Pseudonymisation is gaining traction among modern electronic identification systems as a privacy enhancing technique that can significantly reduce risks of personal data misuse. The recently agreed General Data Protection Regulation (the GDPR) encourages the use of pseudonymisation to comply with its requirement of privacy-by-design. Art. 5 of the European Regulation on electronic identification and trust services (eIDAS) on data processing and protection simply allows the use of pseudonyms in electronic transactions although facilitating the implementation of the principle of privacy by design is clearly among the aims listed by Art. 12 of eIDAS. This paper examines the concept of pseudonymisation under eIDAS and the GDPR and suggests that the two Regulations employ two very different, if not unrelated, notions. It concludes that a common terminology and approach would be preferable in order to ensure consistency and legal certainty.

Keywords: eIDAS, the GDPR, pseudonymisation, e-ID, electronic identity, eIDM

1 Introduction

This paper focuses on pseudonymisation, a concept that was only recently formally introduced in the EU regulatory landscape. In particular it attempts to derive the effects of the introduction of pseudonyms (or pseudonymous credentials) as part of the Regulation on electronic identification and trust services (eIDAS) and ultimately to compare them with the effects of pseudonymisation within the meaning of the General Data Protection Regulation (the GDPR). The paper thus examines how eIDAS conceives pseudonymisation and explains how this interpretation would translate in practical uses in the context of a pan-European interoperability framework. It then assesses this approach in the light of the newly adopted data protection rules, i.e. the final version of the GDPR. the GDPR, as adopted on the 27th of April 2016, introduces a new category of data, between personal and data that has undergone anonymisation, that seems to ease the further processing of personal data beyond the initial processing purposes. Finally the paper concludes with some useful insights about the intersection of the two Regulations and possible inconsistencies.

The paper is organised as follows: Sect. 2 offers an overview of eIDAS Regulation and its requirements in sect. 2.1. Sect. 2.2 examines how eIDAS conceives pseudonymisation
and a parallel is drawn with the definition and approach of the GDPR in sect. 3. Sect. 3 combines both eIDAS and the GDPR to assess the meaningfulness of the requirements of the GDPR. Finally, sect. 4 concludes by stressing the need to redress how pseudonymisation is currently defined in the GDPR and by questioning the way privacy by design is promoted in eIDAS in the light of the GDPR.

2 eIDAS Regulation

In 2012 the Commission decided that a revision to the eSignatures Directive was necessary. The proposed draft was adopted by the Parliament and the Council in 2014. eIDAS’ aim is to create a uniform legal framework under which the various national electronic identification schemes and trust services will be able to interoperate in cross-border transactions. In line with most recent policy initiatives, eIDAS follows a technology-agnostic approach in regulating the interoperability framework. The Regulation itself sets out the goals of the policy but does not define any system specifications. According to eIDAS member states that wish to operate cross-border authentication should notify their national eID Management systems (eIDM) to a designated body. Voluntary notification applications are accepted since September 2015 and by 2018 all member states are required to start accepting identifications from notified schemes.

2.1 eIDAS minimum requirements

As alluded above, the Regulation and implementing acts try to keep requirements to a minimum. eIDAS defines three Levels of Assurance that define a risk-based identification, requiring higher levels of identification authentication depending on the importance of the transaction about to be performed. The levels are named ‘Low’, ‘Substantial’ and ‘High’ and their requirements are defined through an implementing act. A specific mention is made to data protection, as all components of the framework are required to fully comply with the Data Protection Directive (DPD), which was in effect when the Regulation was published. The framework shall also facilitate the ‘Privacy-by-design’ principle. 

---

6eIDAS n. 5 Arts. 6, 7 and 9.
7Specific technologies are acceptable only when absolutely essential for the security of the system or the users. Note that the first drafts of the Regulation were even stricter, prohibiting member states to operate any scheme that would require extra hardware or software to be implemented by other member states. This wording was toned down in the final text after objections: See [CS14]. However, the implementing acts point towards specific implementations, creating, therefore, de facto standards.
10eIDAS n. 5 Art. 12(3d).
11eIDAS n. 5 Art. 12(3c). Article 12(3c) provides that the interoperability framework shall (so it is a mandatory requirement) facilitate the implementation of the principle of privacy by design. What the word facilitate implies is obviously not entirely clear, but Article 12(3c) shall now also be read together with Article 25 of the GDPR.
Though strictly speaking not legally binding, the recitals state that online services should comply with the data minimisation principle and request and process only data strictly necessary for each transaction.12 The goal of the interoperability framework is to make identification ‘uniquely representing’ a natural or legal person possible within a cross-border context.13 Implementing Regulation 2015/1501 clarifies this further by providing a minimum dataset (MDS) that shall be transmitted for every identification.14 The MDS defines four mandatory and four optional identifiers that need to be included (table 1). Note that the presence of a unique identifier is mandatory. Member States can decide which identifier to use,15 as long as the identifier is as persistent in time as possible. As a result, the interoperability framework seems to allow only limited selective disclosure16 insofar as the user can decide to enrich or not the minimum dataset with additional attributes.17 The decision of the regulating committee to introduce the minimum dataset has been questioned for designing a framework that offers less privacy than what is currently technically possible [MG13, AB14, ZS13, p. 6].18

2.2 Pseudonyms under eIDAS

Art. 5(2) of eIDAS allows the use of pseudonyms for electronic transactions. Although Art. 5(2) does not expressly mention electronic identification one could try to make the which states, among other things, that the “The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed”.

12eIDAS n. 5 recital 11.
13eIDAS n. 5 Art. 3(1).
15According to the published SAML profile, the identifier will be composed out of a standard alpha-2 code for the State of origin, a standard alpha-2 code for the State of destination and a combination of characters that “uniquely identifies the identity asserted in the country of origin but does not necessarily reveal any discernible correspondence with the subject’s actual identifier (for example, username, fiscal number etc.)”: [Ei15, p. 8].
16The concept of selective disclosure allows the revealing of only the parts of the information available in an eID necessary for a particular transaction. For example, when only proof of the current address is required, selective disclosure provides only this information to the relying party [PH10].
17This seems to be a ‘lighter’ version of full selective disclosure and certainly of the German understanding of selective disclosure. The German nPA allows the user to control any attribute sent in a transaction (the service provider has the right to refuse a transaction if not all the necessary attributes for the transaction have been disclosed however) [Fe11, pp. 24–25].
18Please note that the papers referred here were published before the final text of eIDAS; some of the arguments have since been mitigated, for example the final text of eIDAS is less strict on hardware and software requirements in Rec. 19 “appropriate solutions should be discussed and developed within the scope of the interoperability framework. Nevertheless, technical requirements stemming from the inherent specifications of national electronic identification means and likely to affect the holders of such electronic means (e.g. smartcards), are unavoidable”; cf. COM/2012/238 where under Rec. 15: interoperability “rules out any specific national technical rules requiring non-national parties for instance to obtain specific hardware or software to verify and validate the notified electronic identification”). The arguments concerning the MDS, though, should still be considered to hold true insofar as the framework cannot offer selective disclosure (at least for the attributes included in the mandatory section of the MDS), disallowing, therefore, the use of 0-knowledge authentication technologies (Cf. though [KKA14] arguing that such technologies might result in stricter authentication if widely implemented ); in the case that the framework does not support selective disclosure of attributes and the UID has to be transmitted in every transaction, a question of potential linkability issues across uses could be raised.
argument that because Art. 5(2) is contained in the general part of eIDAS it ought to apply to both electronic identification and trust services.

"Without prejudice to the legal effect given to pseudonyms under national law, the use of pseudonyms in electronic transactions shall not be prohibited."19

Pseudonyms are artificial identifiers constructed to replace some information in a dataset in order to make attribution of a record to a person harder.20 Pseudonyms are not included in the definitions of eIDAS and the only other mentions in the Regulation are in relation to qualified certificates. This raises the question whether the inclusion of pseudonyms in Art. 5(2) was only meant to be of relevance for qualified certifications or whether it could be seen as an attempt to encourage the use of privacy-enhancing technologies in the context of national eID systems (eIDM). Pseudonyms offer better privacy protection [PH10, Br14, Ve15] and have been used by national eID already to enhance user privacy. In Austria, eIDs include a Unique Identifier (UID) that derives from the Central Residents Register. The architecture resembles in some respects Estonia’s, which also uses central Resident Numbers. In contrast to Estonia though, where Residents Numbers are publicly available information, in Austria it is prohibited by law to share this number with the services.21 Instead, the system employs sector-specific PINs (ssPin) – pseudonyms constructed partly from citizen number and partly from the services requiring them [Rö08]. In this way every department receives

<table>
<thead>
<tr>
<th>mandatory</th>
<th>optional</th>
</tr>
</thead>
<tbody>
<tr>
<td>current family name(s)</td>
<td>first and family name(s) at birth</td>
</tr>
<tr>
<td>current first name(s)</td>
<td>place of birth</td>
</tr>
<tr>
<td>date of birth</td>
<td>current address</td>
</tr>
<tr>
<td>unique identifier</td>
<td>gender</td>
</tr>
</tbody>
</table>

Tab. 1: Minimum dataset for a natural person

---

19 The Article targets ‘electronic transactions’. Electronic transactions are not included in the list of definitions of Art. 3. Art. 1, dealing with the subject matter of the Regulation, refers to electronic transactions in the context of trust services “...[this Regulation] lays down rules for trust services, in particular for electronic transactions” (Art. 1(b)). However, according to Recital 2 eIDAS aims to enhance trust in electronic transactions by providing “a common foundation for secure electronic interaction between citizens, businesses and public authorities.” In addition Recital 17 asks Member States to encourage the use of electronic identification means “when needed for online services or electronic transactions.” It could seem reasonable to conclude that, in absence of an explicit exclusion, Art. 5 refers to both eID means and trust services. Such an interpretation would be in line with the placement of the article within the General Provisions Chapter of the Regulation, which should apply to all following Chapters unless exceptions are carved out.

20 [PH10] defines pseudonyms as “an identifier of a data subject other than one of the subject’s real names.” From a technical point of view pseudonyms are another kind of attribute values associated with a data subject. As such, using pseudonyms to achieve non-attribution of data to a data subject depends on the particular setting where pseudonyms are used. A distinction should be made between randomly and non-randomly produced identifiers: A pseudonym generated by random data only, i.e. fully independent of the subject’s attribute values, do not contain side information about the subject. In contrast, non-random pseudonyms may do, for example a sequence number containing information on the time the pseudonym was issued or an email address in place of the subject’s name, containing information on how to reach the user.

different credentials for the same user and the eID is cross-sector unlinkable which makes it difficult to trace the transactions of a user across the system, i.e., the eIDM does not allow the tracing of a user’s transactions across services. Pseudonyms are employed by the German nPA as well. The German system revolves around the eID card, the nPA, since there is no Identity Provider present. Instead, the user is in charge of their eID and decides when it will be disclosed. Every time nPA is used for a service a specific pseudonym is created that is unique to this particular use [Po12]. Pseudonymisation as a requirement is actually found in various pieces of German legislation. 22

The use of pseudonyms in the context of electronic identification as regulated by eIDAS could be seen as a way to support eIDMs inspired by the Austrian or German model. This interpretation could appear reasonable given that pseudonyms are mentioned immediately after the general requirements to process personal data in accordance with the DPD as per the first paragraph of Art. 5. Although the DPD does not expressly mention pseudonyms or pseudonymisation as a process to ensure compliance, the GDPR, which will soon replace the DPD, introduces a new definition, that of pseudonymisation.

3 Combining eIDAS and the GDPR

Recently the Council, Commission, and Parliament agreed on the final text of the GDPR. 23 The GDPR is set to replace the existing DPD on 25th of May 2018. As this is a Regulation, it will have direct effect within all Member States.

The GDPR includes for the first time a definition for pseudonymisation:

‘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;

The additional information needs to be kept separately by the data controller, who must take all appropriate technical and organisational measures to ensure non-attribution. 24 Pseudonymised data are not exempt from the GDPR. Although recital 28 acknowledges that

---

22 For example, see the ‘telemedia act’ that mandates users use and pay for services pseudonymously: Telemedien gesetz vom 26. Februar 2007 (BGBl. I S. 179). See also the Constitutional Court decision that bans the use of any UID in Germany: Volkszählung Urteil des Ersten Senats vom 15. Dezember 1983 auf die mündliche Verhandlung vom 18. und 19. Oktober 1983, BVerfGE 65, 1, 1 BvR 209, 269, 362, 420, 440, 484/83 in den Verfahren über die Verfassungsbeschwerden [in German].


24 ibid. Art. 4(5).

25 Another innovation of the GDPR is the introduction of a duty to ensure data-protection-by-design and by default: ibid. Art. 25.
pseudonymisation can reduce risks of personal data breaches, under recital 26 pseudonymised data should still be considered as personal data inasmuch as they include information relating to identifiable natural persons.\textsuperscript{26}

Unfortunately, the GDPR’s definition of pseudonymisation seems to be too strict to be practically attainable. And, even though ‘pseudonymisation’ is expressly defined, not once a reference to what a pseudonym is given.

In order for any dataset to be considered pseudonymised within the meaning of the GDPR it shall not be possible to attribute information to identifiable individuals. This would thus mean that as long as selective disclosure is not permitted it would be impossible to speak about pseudonymised data in the context of electronic identification. Besides, even if selective disclosure is an option, it is not certain whether it would be possible to characterise data that has undergone pseudonymisation as to determine whether an individual is identifiable. Account should be taken of the means of the data controller and the means of third parties. As aforementioned, selective disclosure allows the system to transmit only those attributes of the eID that are absolutely necessary for the needs of each service \cite{PH10}, i.e. if a service needs only to know if a user is a citizen or not, the system can reply with only a pseudonym for the particular user and a Yes/No answer to their citizen status. This is a function that is included in modern privacy-preserving systems, such as in the German nPA \cite{Fe11}. The nPA could potentially transmit completely pseudonymous datasets in the sense that the use of pseudonyms coupled with selective disclosure makes re-identification more difficult. Whether one would then be able to describe the process as pseudonymisation within the GDPR is not entirely sure. For example, if a service only requires a user to be over a certain age, the system will send only a pseudonym for this use \textsuperscript{27} and a Yes/No answer based on a calculation of the user’s age according to their date of birth. In this case, the chance of re-identifying the person based on indirect identifiers could seem to be minimised. On the other hand, in a dataset where the pseudonym is accompanied by a date of birth, address and gender the possibility of re-identification is significantly higher.\textsuperscript{28}

As aforementioned and as required by Implementing Regulation 2015/1501, it is mandatory for the MDS to \textit{always} include all of the identifiers of the mandatory section. It seems, therefore, that pseudonyms were never conceived as a privacy-enhancing tool for electronic identification purposes,\textsuperscript{29} meaning that the interoperability framework could appear as offering less privacy protection than the individual national systems are capable of.\textsuperscript{30}

\textsuperscript{26}The GDPR seems to propose a risk-based analysis to delineate its material scope, based on the test of the means reasonably likely to be used by both the data controller and third parties. Such a test has been criticised by some commentators for being too subjective \cite{Es16}.

\textsuperscript{27}The German system is capable of performing calculations based on the date of birth. Every identifier of an eID resides inside the eID card of the user and the system reads and transmits only the necessary information for each use \cite{Po12}.

\textsuperscript{28}According to \cite{Sw00} these three indirect identifiers are enough to identify 87\% of the American population.

\textsuperscript{29}See, for example, \cite{ZS13} where the authors explain how pseudonyms and selective disclosure can enhance system privacy.

\textsuperscript{30}At least in so far as attribute-based selective disclosure is not possible for the Minimum DataSet. See n. 18 above.
Such a choice could be explained by the necessity to accommodate all available national architectures. Detailing the differences in eIDM architecture is beyond the scope of this paper, but implementation varies from systems with central governmental databases that use the Central Residents Register number as a UID, like Estonia’s [Ma10], to systems employing a varying UID like Austria [Ro08] and systems designed specifically to disallow the existence of such UID, like Germany’s nPA [Po12] and UK’s Verify [BD11].

While requiring a mandatory UID at the national level would have prevented some member states from notifying their schemes, the use of pseudonyms was not the norm at the time of the adoption of eIDAS. However, pseudonyms have been integrated into the design of the nPA interoperability functionality [BS15] and they have been proposed as a necessary architectural amendment of UK’s Verify [Ts16]. Such an evolution seems compatible with Art. 87 of the GDPR which appears to acknowledge that member states are free to choose any identifier in place of a national identification number. Notably, though, substituting a national unique identifier by a pseudonym will not impact upon the dataset’s characterization: since alongside the pseudonym each dataset has to contain at a minimum the rest of the mandatory identifiers, eIDAS datasets can never satisfy the GDPR’s definition of pseudonymised data. As mentioned in n. 11, Art. 25 the GDPR imposes certain duties on data controllers to implement technical and organisational measures, including pseudonymisation. Since the GDPR is applicable to eIDAS as a whole (through Art. 5(1)), Art. 25 of the GDPR applies to eID schemes. Such measures (or lack thereof) should be justified in the light of the state-of-the-art, the context and purposes of processing and the cost of implementation. The question is thus whether the choice not to promote full selective disclosure could be justified in particular by taking into account the cost of implementation as well as the nature, scope, context and purposes of processing. Such a questioning should not be disregarded too quickly as Art. 83(4) of the GDPR mandates that infringement of Art. 25 is “subject to administrative fines up to 10 000 000 EUR, or in the case of an undertaking, up to 2% of the total worldwide annual turnover of the preceding financial year, whichever is higher.”

31Note that unique identifiers with a national-wide application do not necessarily have to be National Unique Identifiers (in the form, for example, of Austria’s Central Residents Register). Unique identifiers of individual systems could be regarded as de facto national unique identifiers if they gain a wide use; such is the case, for example, in the US where the Social Security Number is being used across finance, employment and governmental agencies (see for example Michigan Department of State v. United States, 166 F. Supp. 2d 1228 on its use when applying for a driving license). Consequently, a unique number that links a citizen’s eID with public-sector records about their activity could potentially be regarded as a de facto National Unique Identifier.

32“Member States may further determine the specific conditions for the processing of a national identification number or any other identifier of general application”: the GDPR n. 23 Art. 87.

33“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” [our emphasis] the GDPR n. 23 Art. 25(1).
4 Conclusion

To conclude it seems that eIDAS and the GDPR do not have the same approach to pseudonymisation. The GDPR promotes pseudonymisation as a useful compliance tool to mitigate personal data risks and effect data minimisation and data-protection-by-default [Es15]. Yet, its proposed definition appears over-restrictive [SK16] and will create confusion between pseudonymous datasets (that still fall within the scope of the GDPR) and anonymous datasets (that are exempted). On the other hand, eIDAS could seem to be regarding pseudonyms simply as a means to create unique identifiers, failing to incorporate pseudonyms in a way that would allow the system to enjoy full selective disclosure functionality – a fact that was highlighted during the drafting phase of eIDAS by the Article 29 Working Party [Ar15]. In other words, pseudonyms would not be conceived as a means to reduce linkability assuming they can be used in relation to electronic identification. This raises the question whether the GDPR and its Art. 25 could not potentially be more prescriptive than that.

Going further, in a recent communication about the future of online platforms, the Commission expresses the opinion that online platforms should be encouraged to start accepting eIDs from schemes notified under eIDAS as a means to authenticate their users [Eu16]. The possibility of using a system that was designed essentially to be of interest in the context of public services could appear to be problematic. This is true in particular if full selective disclosure is not meant to be allowed or promoted for cross-border transactions.

Acknowledgements

This research was funded by the Research Councils UK Digital Economy Programme, Web Science Doctoral Training Centre, University of Southampton, EP/L016117/1, partly supported by the SOCIAM Project, funded by the UK Engineering and Physical Sciences Research Council under grant number EP/J017728/2, and was partly funded the European Union’s Horizon 2020 research and innovation programme under grant agreement No 700542. This paper reflects only the authors’ views; the Commission is not responsible for any use that may be made of the information it contains.

The authors would like to thank @sam280 for the useful discussion that helped to focus some sections of this article.

Literature


34 “In order to empower consumers and to safeguard principles of competition, consumer protection and data protection, “the Commission will further promote interoperability actions, including through issuing principles and guidance on eID interoperability at the latest by 2017. The aim will be to encourage online platforms to recognise other eID means – in particular those notified under the eIDAS Regulation” [Eu16, p. 11] [emphasis in original].


Esayas, S. Y.: The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the ‘all or nothing’ approach. European Journal of Law and Technology, 6(2), 2015.


