Anonymization Between Minimization and Erasure: The Perspectives of French and Italian Data Protection Authorities

Anonymization Between Minimization and Erasure: The Perspectives of French and Italian Data Protection Authorities

Emanuela Podda and Francesco Vigna

Alma Mater Studiorum, CIRSFID - Alma AI, Università Di Bologna, Bologna, Italy
{emanuela.podda,francesco.vigna}@studio.unibo.it

Abstract. Two years after the General Data Protection Regulation (GDPR) went into effect, data anonymization remains one of the main issues linked to fragmentation in the Member States' anonymization policies, in which regard stakeholders would like additional guidelines.

In keeping with this premise, this article aims to analyze and compare trends in the implementation and enforcement of anonymization policies put in place by data protection authorities in two countries: France and Italy.

This analysis makes it possible to trace the evolution of these policies and highlight their critical importance in applying the data minimization principle and in enforcing the right to erasure under Art. 17 GDPR.

Keywords: Anonymization · Risk-based approach · Data protection · Data minimization · Right to erasure

1 Introduction

The anonymization of personal data, stemming from the statistical context, was originally used by the National Statistical Institutes to ensure a sustainable trade-off between data utility and data protection. With the advent of big data it has been extensively applied to *real data*, posing remarkable challenges and tensions in terms of data protection and privacy.

To overcome and relieve this tension, the General Data Protection Regulation (GDPR) seems to be relying on the right of data subjects to be informed about the way data is processed and on the risk-based approach.

Anonymization generally protects data and takes data outside of the scope of data protection and privacy: information rights ensure that data subjects have control over their own data, and risk assessment obliges data controllers to evaluate the level of the risk inherent in the data processing.

The level of anonymity should consequently be tailored to the needs agreed to during the information-giving process, as a person's identity can be revealed by way of either.

– direct identifiers, such as names, postcodes, or pictures; or

– indirect identifiers, which do not in themselves identify anyone but *can* do so in combination with other information available about them, examples of such indirect identifiers being information about someone's workplace, occupation, salary, or age.

For this reason anonymization, together with pseudonymization, represents the main data processing tool the law provides for protecting data and personal identity.

For an official definition of anonymization in the legal data-protection framework in force in Europe, the only source we can turn to is Opinion 05/2014 on Anonymization Techniques[1], issued by Working Party 29, in which anonymization is defined as "*the process by which data are made anonymous*". So defined, anonymization means that the data in question must be stripped of identifying elements, enough so that the data subject can no longer be identified, which in turn means that in order for the data to count as anonymous, anonymisation must be irreversible under all conditions.

From a mere legal point of view, the friction point lies in what is considered an acceptable level of reidentification risk: the legislator acknowledges the problem of reverse-engineering the anonymization process and circumventing data protection tools—a problem that will continue to persist as long as the available technology and technological development keep advancing.

Indeed Recital 26 of the GDPR stresses that "*[t]he principles of data protection should apply to any information concerning an identified or identifiable natural person.* […]" [2].

The whole structure of this Recital strengthens the approach contained in the WP29 Opinion on Anonymization Techniques, where the main focus is not on anonymization *per se* but rather on its outcome. Indeed, in this opinion Working Party 29 stresses the need to test the anonymization techniques against three main risks:

– singling out
– linkability
– inference.

---

[1] See https://ec.europa.eu/info/law/law-topic/data-protection/communication-two-years-application-general-data-protection-regulation_en. The Communication is also accompanied by a Staff Working Document detailing the main issues relating to fragmentation in the Member States' policies and providing guidance for the follow-up. See https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020SC0115&from=EN.

[2] The article continues as follows: "*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. 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. 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 or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes*".

Consistent with this background is the advice provided by WP29: the optimal solution for crafting an anonymization process needs to be decided on a case-by-case basis, as it should be aimed at reducing the risk of deanonymization.

The literature on this topic is vast but can be broken down into two main buckets: on one hand is research focused on empirical cases of deanonymization[3]; on the other the research asks whether compliance requires a zero-risk approach or whether a residual-risk approach will do[4]. Therefore, it seems that the discussion on anonymization mainly revolves around the question of risk (zero vs. acceptable risk), without offering any theoretical approach on which to implement and provide solutions for overcoming this impasse.

[3] In 1990, it was demonstrated that the governor of Massachusetts could be reidentified from deidentified medical data by cross-referencing the deidentified information with publicly available census data used to identify patients. See Sweeney, L.: Policy and Law: Identifiability of de-identified data, in http://latanyasweeney.org/work/identifiability.html. In 2006, the online service provider AOL shared deidentified search data as part of a research initiative, and researchers were able to link search queries to the individuals behind them. See https://en.wikipedia.org/wiki/AOL_search_data_leak. In 2009, Netflix released an anonymized movie rating dataset as part of a contest, and researchers successfully reidentified the users. See Narayanan, A., Shmatikov, V.: Robust De-anonymization of Large Sparse Datasets in https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf. In 2013, another study, conducted on anonymized cell phone data, showed that: "*four spatio-temporal points are enough to uniquely identify 95% of the individuals*". See de Montjoye, Y., et al.: Unique in the Crowd: The privacy bounds of human mobility, Scientific Reports volume 3, Article number: 1376 (2013), https://www.nature.com/articles/srep01376. In 2017, researchers released a study stating that "*web-browsing histories can be linked to social media profiles using only publicly available data*". See Su, J., et al.: De-anonymizing Web Browsing Data with Social Networks, in Proceedings of the 26th International Conference on World Wide WebApril 2017 Pages 1261–1269, in https://doi.org/10.1145/3038912.3052714. Recently, studies have shown that deidentified data was in fact reidentifiable, and researchers at UCL in Belgium and Imperial College London found that "*99.98% of Americans would be correctly re-identified in any dataset using 15 demographic attributes*". See Rocher, L., Hendrickx, J. M., de Montjoye, Y.: Estimating the success of re-identifications in incomplete datasets using generative models, in Nature Communications, Volume 10, 3069, 2019, https://ui.adsabs.harvard.edu/abs/2019NatCo..10.3069R/abstract. The amount of personal information leaked keeps growing, as the exposition of personal data through breaches keeps increasing.

[4] Spindler, G., Schmechel, P.: Personal Data and Encryption in the European General Data Protection Regulation. Journal of Intellectual Property Information, Technology & Electronic Communication, 163 (2016); Kuner, C., et al.: Risk management in data protection. International Data Privacy Law, 2015, Vol. 5, No. 2.; Veale, M., Binns, R., Ausloos, J.: When data protection by design and data subject rights clash. International Data Privacy Law, 2018, Vol. 8, No. 2; 7. Gellert, R.: We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences between the Rights-Based and the Risk-Based Approaches to Data Protection. European Data Protection Law Review, 481 (2016); Gellert, R.: Understanding the notion of risk in the General Data Protection Regulation. Computer Law & Security Review 34 (2018) 279–288; Finck, M., Pallas, F.: They who must not be identified—distinguishing personal from non-personal data under the GDPR. International Data Privacy Law,2020, Vol. 10, No. 1.

Moreover, these polarized perspectives and approaches have certainly come to the notice of investors and stockholders. Indeed, when the European Commission, pursuant to Art. 97 GDPR[5], released its first report on the first two years of GDPR application[6], it listed anonymization among several areas for future improvement. The Commission's report highlights a certain fragmentation in the policy landscape across Member States and data protection authorities, while stakeholders are asking for additional guidelines from the European Data Protection Board (EDPB)[7]. It turns out that there is a great need for theoretical solutions by which to implement anonymization in complying with the principle of data minimization (*ex-ante processing*) and the principle of storage limitation (*ex-post processing*), and even in enforcing Art. 17 GDPR (the right to erasure, or right to be forgotten).

One of the main guidelines provided by the EDPB can be recalled in Opinion 04/2019, addressing the effectiveness of anonymization/deletion in light of these principles. The EDPB specified that whenever personal data is no longer needed after it is first processed, *it must by default be deleted or anonymized*. In keeping with this premise, the EDPB specified that any retention should be objectively justifiable and demonstrable by the data controller in an accountable way, thereby bearing out the view that anonymization of personal data is an *alternative to deletion*, provided that all the relevant contextual elements are taken into account and the likelihood and severity of the risk, including the risk of reidentification, is regularly assessed.

In both cases (deletion and anonymization) the controller is obligated to limit the retention period to what is strictly necessary.

A survey of all the different approaches followed by data protection authorities (DPAs) will reveal that some of them gain paramount importance because they seem to set the trend in clearing up these gray areas of interpretation and practice.

---

[5] Art. 97 GDPR requires the Commission to review the regulation, issuing an initial report two years after its entry into force and every four years thereafter.

[6] Communication from the Commission to the European Parliament and the Council - two years of application of the General Data Protection Regulation, 24 June 2020, Justice and Consumers, COM(2020) 264 final; SWD(2020) 115 final & Staff Working Document accompanying the Communication - two years of application of the General Data Protection Regulation. See https://ec.europa.eu/info/law/law-topic/data-protection/communication-two-years-application-general-data-protection-regulation_en.

[7] The Commission has also issued a study titled "Assessment of the Member States' Rules on Health Data in the Light of GDPR" (Specific Contract No SC 2019 70 02 in the context of the Single Framework Contract Chafea/2018/Health/03), in which anonymization is recognized as an important issue among Member States. See https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf.

This paper therefore intends to analyze and compare the approaches taken by two of the DPAs that are leading the way when it comes to policy implementation, data reuse, open data approach, and enforcement: the French and Italian DPAs[8].

2 The Approach of the French Data Protection Authority: The Commission Nationale de l'Informatique et des Libertés

The French Data Protection Authority—the Commission Nationale de l'Informatique et des Libertés (or CNIL)—defines personal data as *any anonymous data that can be double-checked to identify a specific individual*. Under this definition, certainly stricter than the one contained in Art. 4 GDPR, the CNIL is expressively stating that personal data can be deemed personal even after it has been anonymized.

The CNIL highlights that the process of *anonymizing process personal data should make it impossible to identify individuals within data sets*, and that this must therefore be an *irreversible process.* It follows that *when anonymization is effective, the data are no longer considered personal and the GDPR is no longer applicable.* However, the CNIL recommends that *raw datasets* containing personal data which have been anonymized *should never be deemed anonymous*, requiring a case-by-case evaluation in keeping with the WP29 opinion.

If we look at the trendline in the case-by-case approach adopted by the CNIL, we will find that since the WP29 Opinion on Anonymization Techniques was released, the approach was making possible to store personal data (after the time limit, consistently with the purpose of collection) for statistical purposes, once *effectively* anonymized—thereby confirming the storage exception allowed for statistical purposes under the GDPR.

However, apart from the statistical exception, the approach adopted by the CNIL was also based on detecting the risk of deanonymization, specifying that a dataset is *effectively* anonymous if, prior to dissemination, it meets the following criteria:

- It does not contain individual data.
- It is not possible to correlate the data with any other dataset.
- It is not possible to infer any new information about individuals.

If, and only if, all three criteria are simultaneously satisfied can the dataset be considered *anonymous*, meaning that if even only one criterion is not satisfied, the CNIL will

[8] In December 2020, the European Data Portal issued an annual benchmark study on developments in open data, listing France among the trendsetters and Italy among the fast-trackers. See https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf.Moreover, Italy and France are listed among the top ten Countries with the highest fines in GDPR enforcement. See https://www.enforcementtracker.com/ In details, see also: Daigle, B., & Khan, M.: The EU General Data Protection Regulation: An Analysis of Enforcement Trends by EU Data Protection Authorities. Journal of International Commerce and Economics. May 2020. https://www.usitc.gov/joumals.

have to carry out a specific impact assessment on the risk of reidentification, demonstrating that the dissemination of the anonymized dataset has no impact on the data subjects' private lives or freedoms[9].

The trend that can be observed in the CNIL's practice in the wake of WP29 Opinion 05/2014 shows a remarkable number of authorizations for anonymization processing requested by firms, suggesting that firms want to avoid taking the deanonymization risk, and would much rather take the safer and more convenient route of falling back on the CNIL to test anonymization techniques[10].

Given the risk of deanonymization, and especially the indirect risk due to data export, the CNIL, in authorizing anonymization processing, reiterates that only anonymous data can be exported outside an *approved environment* (especially when it comes to health and research data, for which the rules are certainly stricter), confirming that personal data export is prohibited.

Only *anonymous* data can be exported, only once internal validation is obtained, followed by a systematic and preliminary audit to attest to the effective anonymity of the requested export, thus imposing a double check on the dataset and confirming that in the risk of deanonymization lies the difference between anonymized and anonymous data. The first risk test is designed to ascertain whether data have undergone an anonymization process, while only the second test ascertains that the risk of deanonymization equals zero[11]. On more than one occasion, the CNIL has stressed that the criteria and procedures should be regularly reviewed in light of changing anonymization and reidentification techniques.

This trend confirms what is started in the CNIL's guidance on anonymization[12], namely, that anonymization processing is not required under the GDPR but is a way the GDPR is implemented under certain conditions.

Along the same lines, personal data need to be treated in secure environments (*bulle sécurisée*) and be stored and processed as long as is necessary for the purposes of collection, after which the data needs to be destroyed.

Moreover, a specific role in the implementation of anonymization is provided for by *Loi 2004–801 du 6 août 2004*, on the protection of individuals in relation to personal data

[9] Among the first cases brought before the CNIL immediately after the publication of the WP29 Opinion on Anonymization Techniques was *Délibération N° 2015–425 du 3 décembre 2015 autorisant l'association réseau périnatal de l'Ile-de-France Sud à mettre en œuvre un traitement automatisé de données à caractère personnel dénommé « Hygie TIU» ayant pour finalité le suivi des transferts in utero en Ile-de-France.*

[10] Among the most recent examples, see *Délibération n° 2020–055 du 14 mai 2020; Délibération n° 2019–124 du 10 octobre 2019; Délibération n° 2019–112 du 5 septembre 2019; Délibération n° 2019–122 du 3 octobre 2019; Délibération n° 2019–110 du 05 septembre 2019*; etc.

[11] An extensive definition can be found in *Délibération N° 2015–425 du 3 décembre 2015.*

[12] *CNIL, L'anonymisation de données personnelles, 19 mai 2020*, cfr. https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles.

processing[13], amending *Loi n° 78–17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés* (Law 78–17 of January 6, 1978, relating to computers, records, and freedoms).

The 2004 law, anticipating the new framework proposed by the European Commission in its first draft of the Data Governance Act[14], states that the CNIL may certify and publish general repositories or methods for certifying anonymization through the services of approved or accredited third parties.

What initially emerges from the foregoing analysis is that the French DPA took a more stringent attitude towards anonymization, relying on specific and multiple tests and controls for detecting and ensuring that the risk of deanonymization is down to zero.

3 The Approach of the Italian Data Protection Authority (Garante per la Protezione dei Dati Personali)

The Italian Data Protection Authority—Garante per la Protezione dei Dati Personali (GPDP)—has never provided its own interpretation of the concept of anonymization, at least not explicitly: There are to date no opinions or guidelines the GPDP has issued specifically explaining anonymization as a concept. However, the GPDP has dealt with the issue of anonymization several times, in general and special provisions alike.

Instead, the GPDP usually refers to the interpretation given by the Article 29 Working Party (WP29), or, sometimes, even to documents issued by the Information Commissioner Officer (ICO) or by the Commission Nationale de l'Informatique et des Libertés (CNIL)[15].

The problem of how the anonymization concept is to be interpreted has been widely discussed in the literature. Insightful contributions focus on the risk approach for evaluating the robustness of the anonymization technique[16].

But what is the position of the GPDP on evaluating the risk of anonymization techniques? An investigation of the acts issued by the GPDP shows that the Italian authority does not adopt the same approach in all situations. Which is to say that on some occasions the GPDP states that anonymization needs to be irreversible, while on others it

[13] This law was introduced to transpose Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with regard to the Processing of Personal Data and the Free Movement of such Data. It was designed to adapt the law governing computer records to technological advances and contemporary realities, and to do so consistently with the basic principles set forth in law of January 6, 1978. It was in turn amended by *Loi 2018–493 du 20 juin 2018 relative à la protection des données personnelles*, by which the GDPR was transposed into national law.

[14] The 2004 law does so specifically by introducing the concepts of *data altruism* and *secure data-processing environments* and providing for data altruism services in complying with additional technical, administrative, or organisational requirements, including through an authorisation or certification regime. Here, anonymization is framed as a dynamic process using risk analysis, supported by intermediaries and certified actors, and introducing safe environments for data processing.

[15] See, for example, GPDP, doc. web 2014 – 3134436.

[16] In this regard, see note 4 above.

takes account of the risk of reidentification, and on others still it does not indicate at all what approach is to be followed.

For example, in dealing with a number of situations, the GPDP has ordered data controllers to ensure that personal data is "*deleted or made definitively anonymous*"[17], apparently not allowing any residual risk of identifying the data subject. In another provision, using an even stronger *formula*, the GPDP states that personal data—once the retention period has lapsed—must be "*automatically deleted or permanently and irreversibly anonymised*"[18].

Differently, in other measures the GPDP has explicitly recalled the identification risk assessment under Recital 26 GDPR.

For instance, in a document titled "Deontological Data Processing Rules for Statistical or Scientific Research" the authority states that a data subject is deemed identifiable where, "by the use of reasonable means, it is possible to establish a significantly probable association" between a statistical unit and the data identifying the same data subject[19]. The same guidelines also provide some examples of "reasonable means", such as economic resources and time; they also suggest risk-assessment criteria, recommending that data controllers consider how confidential the data at stake is and that they adopt a reasonableness approach in making the assessment[20].

In that same vein, there are several occasions on which the GPDP seems to be aware that the anonymization process cannot be taken as a fixed outcome but is rather something that could change depending on context.

Thus, for example, in regard to the "preliminary check" on data processing carried out for profiling purposes, the GPDP recalls the WP29 opinion on anonymization techniques, stating that reaching a "high degree" of anonymization is a matter of reasonableness. Specifically, the GPDP requires data controllers to take into account all the means that could be reasonably used to identify a data subject, while also, at the same time, evaluating the "likelihood of re-identification"[21] inherent in the data processing.

Finally, there are also situations where the GPDP does not specify whether the approach to be followed should be the irreversible kind or a reasonable-risk assessment, a case in point being where the GPDP merely states that data is to be "deleted or made anonymous"[22], without explicitly taking a stance on the residual risk that the data subject might be reidentified[23].

From a different perspective, the GPDP's acts could be interpreted in light of the purpose ascribed to the anonymization process, meaning that this process may understood as designed to implement either.

---

[17] GPDP, doc web 2020 – 9356568, p. 12.

[18] GPDP, doc web 2018 – 8998319, p. 3; see also GPDP doc web 2018 – 8997404, p. 4.

[19] GPDP, doc web 2018 – 9069637, p. 7; see also, along the same lines, GPDP, doc web 2020 – 9069677.

[20] GPDP doc web 2015 – 4015426; GPDP, doc web 2015 – 3881392.

[21] GPDP doc web 2015 – 4698620, p. 6; see also GPDP, doc web 2020 – 9520567, doc web 2015 – 3843693, doc web 2020 – 9356568 for similar reasoning.

[22] GPDP, doc web 2016 – 4943801, p. 5.

[23] See also GPDP, doc web 2019 – 9124510; GPDP, doc web 2018 – 9068972.

1) the minimization principle or
2) the storage-limitation and the purpose-limitation principles.

These principles have a different role in the data processing: while the first represents the main basis for implementing the security measures, the second acquires importance once the scope of the data collection has been achieved.

In what concerns anonymization as a minimization measure, we might point to an opinion the GPDP issued relating to the Italian contact-tracing app for the COVID-19 outbreak: referring to the EDPB guidelines on scientific research and contact-tracing apps in the COVID-19 context[24], the GPDP reminded the Italian Presidency of Ministers that under the minimization principle, it is necessary to collect "only the data that is strictly necessary to detect possible infections, while using reliable anonymization and pseudonymization techniques"[25].

Along the same lines, the GPDP, in an opinion about a decree scheme put out by the Italian Labour Minister, seems to endorse the draft decree in part because it provides for the use of "anonymous or aggregated data by the Minister […] in keeping the data minimization principle (Art. 5(1)(c) of the Regulation)"[26].

Less recently, but still significantly, in a 2017 provision relating to a particular data processing operation under GPDP scrutiny, the data protection authority stated that the final purposes could be pursued "consistently with the data minimization principle […] without processing the personal data of the clientele […], but only processing anonymized data and/or data for which consent has been obtained"[27].

On the other hand, the Italian DPA requires data controllers to implement the anonymization process as a tool for complying with the principles of purpose and storage limitation; this means that the GPDP allows data controllers to further process personal data without deleting it if data is anonymized[28].

On this approach, the GPDP treats anonymization as equivalent to cancellation. In these cases, an irreversible outcome is usually (but not always) explicitly required by the GPDP[29].

What emerges from this analysis by way of an initial assessment is that the Italian DPA takes a tailored attitude to anonymization, which can become more or less strict depending on the specific situation and even on the data processing context.

Usually, in situations where anonymization is carried out as an alternative to data cancellation, and therefore the scope of data processing is achieved, the GPDP seems to require an irreversible, zero-risk approach. Even if, this last one - as a practical matter, as previously mentioned – cannot be considered to be achieved.

---

[24] EDPB Guidelines 3/2020 and 4/2020.

[25] GPDP, doc web 2020 – 9328050, p. 3.

[26] GPDP, doc web 2019 – 9122428, p. 5.

[27] GPDP, doc web 2017 – 6844421, p. 3.

[28] GPDP, doc web 2020 – 9356568, doc web 2018 – 8998319, doc web 2018 – 8997404, doc web 2015 – 4698620, doc web 2015 – 4015426, doc web 2015 – 3881392.

[29] For cases where irreversibility is not explicitly required even though anonymization is equated with erasure, see GPDP, doc web 2016 – 4943801, doc web 2015 – 4698620.

On the contrary, when the GPDP treats anonymization as an implement measure of the minimization principle (or as a measure for protecting data subjects' rights and freedoms) another approach is requested. Specifically, since the data processing is still ongoing, assessing the re-identification risk is preferable.

4 A Comparative Conclusion

The two approaches followed by the two DPAs certainly bear out the considerations previously made: that there is a great need for theoretical solutions by which to implement anonymization as a tool for complying with the principle of data minimization (*ex-ante processing*) and the principle of storage limitation (*ex-post processing*), and even for enforcing Art. 17 GDPR (right to erasure, or right to be forgotten). Moreover, these approaches show that, overall, anonymization is an essential requirement for data reuse in the digital economy. As such anonymization appears to be called for as an operation to be executed on the datasets which pose a steep challenge on law enforcement.

Looking at the evolution of anonymization in the legal framework and at its follow-up as determined by the previously analyzed trends in the practice of the two DPAs, the WP29 seems to view anonymization as a tool aimed at the *desirable/preferable* zero-risk outcome.

However, over time—at first with the GDPR and national law, and subsequently with the DPAs—anonymization seems to be a tool in constant evolution: the desirable outcome envisioned by the WP29 therefore seems to be becoming utopian.

This certainly cannot mean that its function should be neglected[30], as it remains an essential standard for processing personal data, even if in the risk-management literature it is widely accepted that risk can never be brought to zero[31].

In view of these premises, the question could be whether the GDPR risk-based approach should be considered an exception to the traditional conception of risk (given that the data subject's rights and freedoms are at stake), and so whether a requirement of a zero-risk level should be introduced aimed at protecting fundamental rights and freedoms[32].

Zero risk seems particularly difficult to implement, mainly for two reasons: (1) the concept of risk is by its nature scalable; and (2) a zero-risk level would be impossible to enforce. This last reason is particularly meaningful in the data protection domain, considering how dynamic the context is, especially when innovative technologies are involved.

On the contrary, attention should be focused on the acceptable risk of reidentification, that is, the question should be: When is a risk level low enough to be considered compliant with the data protection legislation? And how to ensure that level?

[30] See, extensively, Biega, A., & Finck, M., Reviving Purpose Limitation and Data Minimisation in Per-sonalisation, Profiling and Decision-Making Systems, Max Planck Institute for Innova-tion and Competition Research Paper No. 21–04.

[31] See, extensively, Gellert, R.: We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences between the Rights-Based and the Risk-Based Approaches to Data Protection. European Data Protection Law Review, 481 (2016).

[32] Gellert, R.: Understanding the notion of risk in the General Data Protection Regulation. Computer Law & Security Review 34 (2018) 279–288.

In this regard, there is an essential role to be played by instruments such as codes of conduct and certification mechanisms (pursuant to Arts. 40–42 GDPR). Even if the GDPR specifies that these solutions can be adopted on a voluntary basis, they could include specific procedures to be followed by data controllers entrusted with compliance. That they can serve as reliable tools of compliance is a view also supported by the mandatory preapproval that needs to be obtained from the national DPA in charge or from the EDBP.

Therefore, on this approach, a twofold objective could be pursued. On the one hand, DPAs could endorse specific ways of applying the GDPR provisions that have no easy interpretation (e.g., the management of the reidentification risk). On the other hand, these instruments would help data controllers in achieving and demonstrating compliance in situations and contexts of particular complexity, ultimately making it possible to standardize the procedures for assessing the risk of reidentification.

Of course, it cannot be left unsaid that, being adopted on a voluntary basis, enforcement could be problematic, even if it can certainly be listed among the *best practices*, setting a responsible trend in mitigating the risks associated with data processing.

In any case, the analysis of the DPAs' provisions also shows that the residual reidentification risk is managed in different ways depending on the purpose of anonymization and its context. In fact, anonymization seems to have different aims[33], since it could be a measure/safeguard for reducing risk, but it could also be construed as further data processing, and it could even be an operation explicitly required by a DPA.

It can even be argued that the optimum level of reidentification risk is ultimately best assessed by looking at the *context* of anonymization. On that basis it would be possible to determine whether anonymization is required by the DPA as an alternative to the cancellation of data, whether the anonymized dataset will be used for purposes different from the original aim of data collection, or whether anonymization is used as measure by which to ensure compliance with the minimization principle.

In conclusion, it appears to the authors of this study that since anonymization in itself constitutes data processing, whenever controllers intend to anonymize personal data, they should do a risk assessment and, if need be, a data protection impact assessment (DPIA).

To this end, in view of the aim of anonymization, particular attention needs to be paid where the output data resulting from the anonymization process are further processed outside the original data context. Indeed, if anonymization is carried out as a data security measure or for data minimization purposes, and then as a measure for decreasing the general data processing level, the risk of reidentification could reasonably be considered lower. On the contrary, if anonymization is used to further process personal data—thereby exiting the original data context and falling outside the scope of the GDPR—the risk of reidentification could be higher.

Overall, a rigorous approach is still desirable in all situations where anonymization allows data controllers to process personal data fully outside the scope of the GDPR and for unrecognized purposes that could jeopardize data subjects' rights and freedoms.

[33] S. Stalla-Bourdillon, A. Knight, Anonymous Data v. Personal Data – a False Debate: an EU Perspective on Anonymization, Pseudonimization and Personal Data.

References

1. Barocas, S., Nissembaum, H.: Big data's end run around anonymity and consent. In: Lane, J., et al. (eds.) Privacy, Big Data and the Public Good, Frameworks for Engagement, Cambridge University Press, Cambridge (2014)
2. Biega, A., Finck, M.: Reviving purpose limitation and data minimisation in personalisation, profiling and decision-making systems. Max Planck Institute for Innovation and Competition Research Paper No. 21-04
3. Daigle, B., Khan, M.: The EU general data protection regulation: an analysis of enforcement trends by eu data protection authorities. J. Int. Commer. Econ. May 2020. https://www.usitc.gov/joumals
4. de Montjoye, Y., et al.: Unique in the crowd: the privacy bounds of human mobility. Sci. Rep. **3**, 1376 (2013). https://www.nature.com/articles/srep01376
5. El Emam, K., Alvarez, C.: A critical appraisal of the Article 29 working party opinion 05/2014 on data anonymization techniques. Int. Data Priv. Law **5**(1), 73–87 (2015)
6. Finck, M., Pallas, F.: They who must not be identified—distinguishing personal from non-personal data under the GDPR. Int. Data Priv. Law **10(**1), 11–36 (2020)
7. Gellert, R.: We have always managed risks in data protection law: understanding the similarities and differences between the rights-based and the risk-based approaches to data protection. Eur. Data Prot. Law Rev. **481** (2016)
8. Gellert, R.: Understanding the notion of risk in the general data protection regulation. Comput. Law Secur. Rev. **34**, 279–288 (2018)
9. Kuner, C., et al.: Risk management in data protection. Int. Data Priv. Law **5**(2), 95–98 (2015)
10. Narayanan, A., Shmatikov, V.: Robust De-anonymization of Large Sparse Datasets in https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf
11. Rocher, L., Hendrickx, J.M., de Montjoye, Y.: Estimating the success of re-identifications in incomplete datasets using generative models. Nat. Commun. **10**, 3069 (2019). https://ui.adsabs.harvard.edu/abs/2019NatCo.10.3069R/abstract
12. Spindler, G., Schmechel, P.: Personal data and encryption in the European general data protection regulation. J. Intellect. Prop. Inf. Technol. Electr. Commun. **163** (2016)
13. Stalla-Bourdillon, S., Knight, A.: Anonymous data v. personal data – a false debate: an EU perspective on anonymization, pseudonimization and personal data. Wiscon. Int. Law J. **34**(284), 284–322 (2017)
14. Su, J., et al.: De-anonymizing web browsing data with social networks. In: Proceedings of the 26th International Conference on World Wide Web April 2017, pp. 1261–1269 (2017). https://doi.org/10.1145/3038912.3052714
15. Sweeney, L.: Policy and law: identifiability of de-identified data. http://latanyasweeney.org/work/identifiability.html
16. Veale, M., Binns, R., Ausloos, J.: When data protection by design and data subject rights clash. Int. Data Priv. Law **8(**2), 105–123 (2018)