DSCI
PROMOTING DATA PROTECTION
A nasscom Initiative

The Future of Data Protection in India

A Roadmap for Regulators

PART 2

First edition: 2023

The Future of Data Protection in India

Published by:

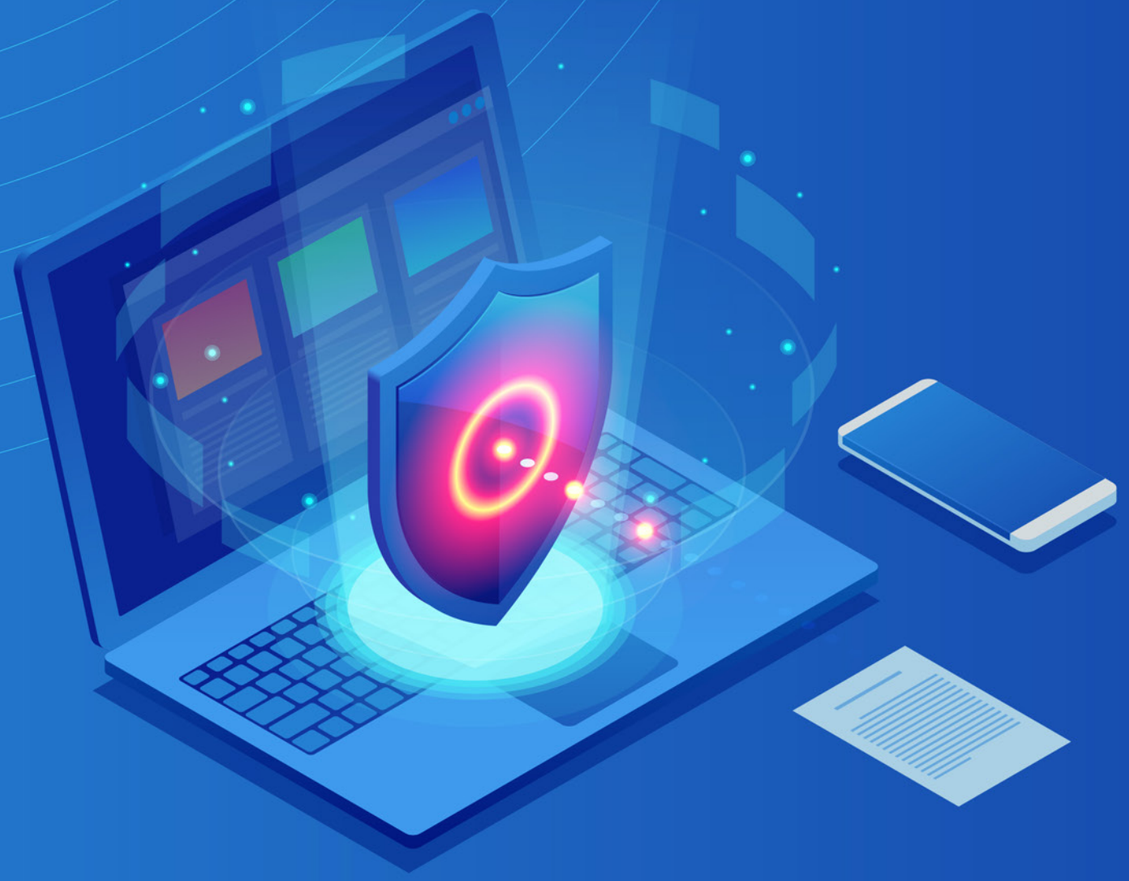
DSCI
PROMOTING DATA PROTECTION
A nasscom Initiative

4th Floor, NASSCOM Campus
Plot No. 7-10, Sector 126, Noida
Uttar Pradesh 201303, India

The Future of Data Protection in India

A Roadmap for Regulators

PART 2



Balancing Privacy and Innovation: Anonymisation Standards for Indian Data

Executive Summary 07

1. Introduction 08
2. Comparing anonymization techniques 11
3. Global perspective: Regulatory guidelines and best practices 13
4. Conclusion 18

Executive Summary

The document seeks to assess and collate best practices for anonymizing data from jurisdictions across the world, taking into account the scope and meaning of anonymization, threshold prescribed, and test of effectiveness. The newly enacted Digital Personal Data Protection Act ('DPDPA') 2023 does not refer to anonymised data, even though it has been touched upon briefly in secondary sectoral policies. The aim of this document is to put forth recommendations for regulatory changes that may be required to be implemented to create an efficient and effective legal regime for implementing and managing anonymization and anonymized data.

The first section delves into the background in which anonymization has gained foothold over the past few decades, alongside a comprehensive exploration of what it may constitute. The link between anonymization and identifiability has been examined in the context of the nature, type and scope of data that may be anonymised in addition to possible use cases. Following this, the section looks at what anonymization techniques are on a graded scale of effectiveness, and where they may be applicable while balancing risk mitigation of identifiability with the utility of emergent dataset. The broad aim of this section is to identify the essentials of anonymization and locate anonymization techniques on the 'spectrum' of identifiability.

The second section focuses on a technical analysis of modern anonymization techniques as against traditional methods for anonymization. It examines the use of k-anonymity and L-diversity to assess the effectiveness of anonymization. The goal of this section is to explore why newer techniques such as homomorphic encryption and differential privacy have been gaining ground, and where primitive techniques of data masking, generalization, swapping and pseudonymization may still find use.

The third section looks at the regulatory approaches to anonymization in the EU, UK, Canada, Singapore, Japan, France, Spain, and Switzerland with the aim of locating the most appropriate practices for the standard of anonymization, governance of anonymization techniques, and risk-mitigation measures.

The final section provides recommendations derived from previous research and analysis. These recommendations are intended for regulatory authorities and policymakers to consider, catering to the needs of industry stakeholders and data principals.

1 Introduction

Anonymization may be understood as a privacy preserving mechanism for data minimization, risk management and data protection in the cases of unauthorised access.

The rise of hyperconnectivity has increasingly limited the ability of organizations and individuals to protect their data. In dealing with personal data, key concerns range from unlawful processing, unauthorized sharing with third parties, inadvertent publication of sensitive information and personal and legal consequences of the same, to intrusive harm, misinformation and even identity theft. The collection and processing of data in the present-day context is all-encompassing, and all organizations pursue real-world personal information for cross-functional integration, better decision-making, and accessing appropriate audiences.

Given the intensity and scale of data collection, it is not always possible for individuals to exercise control over their personal data in a way that meets the privacy expectations of the data subject or for the data fiduciaries to exercise legitimate objectives of processing simultaneously. Data anonymization seeks to address this as a process which removes personally identifiable attributes from personal data reducing or eliminating identifiability.[1] Anonymization, when carried out effectively, is marked by (a) irreversibility, and (b) a strong protection threshold, requiring significant intensive effort to re-identify.[2] The resulting data ceases to be 'personal data', and entities are able to sidestep laborious compliance requirements for personal data, while still meeting their archival and processing purposes. Further, most data protection acts do not apply to anonymised data which cannot be linked to any living person.[3]

ISO Standard 29100: 2011 defines 'data anonymization' as the process by which PII is irreversibly altered in such a manner that the PII Principal can no longer be identified, either by the controller alone or in collaboration with any other party.[4] Anonymization may be understood as a privacy preserving mechanism for data minimization, risk management and data protection in the cases of unauthorised access. Organizations are compelled to use some privacy preserving technique due to either norms, market requirements, technological availability, or law.[5] Firstly, certain contexts such as biomedicine require anonymization by default, and best practice standards for industries such as online commerce suggest it often to maintain privacy, even though the use cases are limited. Secondly, data controllers/ processors are incentivised by market forces to anonymise data for aggregation and sharing to assure customers of their security as well as use their information. Anonymization emerges as one of the techniques which could be implemented, but more economical, trustworthy and practical versions have come up in the recent past. Thirdly, technological constraints allow retention of only a portion of the maximum available personal data and some form of anonymization becomes essential to the interface. Fourthly, it serves as an alternative compliance mechanism to data protection principles in the context of personal data. Anonymization serves the ends of (1) protecting individuals' identities, (2) developing reusable formats of datasets for varied applications, (3) mitigating disputes and freedom of information requests pertaining to processing of personal data, (4) increasing transparency, and (5) forging a public repository of real-world data.[6] However, at the same time, the utility of anonymization is pulled back due to concerns surrounding data ownership, storage, privacy, limited number of use cases, and impossibility of totally anonymized solutions.

> Anonymization techniques include data swapping, data or character masking, attribute suppression, aggregation, pseudonymization, generalization, data perturbation, synthetic data, and encryption.

The process of anonymisation is done through 'anonymization techniques' which may modify, add to, or mirror original datasets to mitigate identifiability of individuals. These techniques vary in terms of the protection offered and fall at different points on the spectrum of identifiability. As such, some of these are more suited to certain contexts depending upon the nature of data, kind of risks posed and intended use. In this context, anonymization standards emerge to ensure effective anonymization of data and provide a broad framework for organizations to operate within. These standards should ideally address the suitability of datasets for anonymization, the test for identifiability, determination and mitigation of risks, governance of anonymized data, and appropriate techniques to be deployed. The aim in this instance is to look at the threshold of anonymization, techniques available in present-day, and the various standards prescribed by regulators around the world to understand how India can forge a path ahead in the arena of anonymization.

1.1 When is data anonymised?

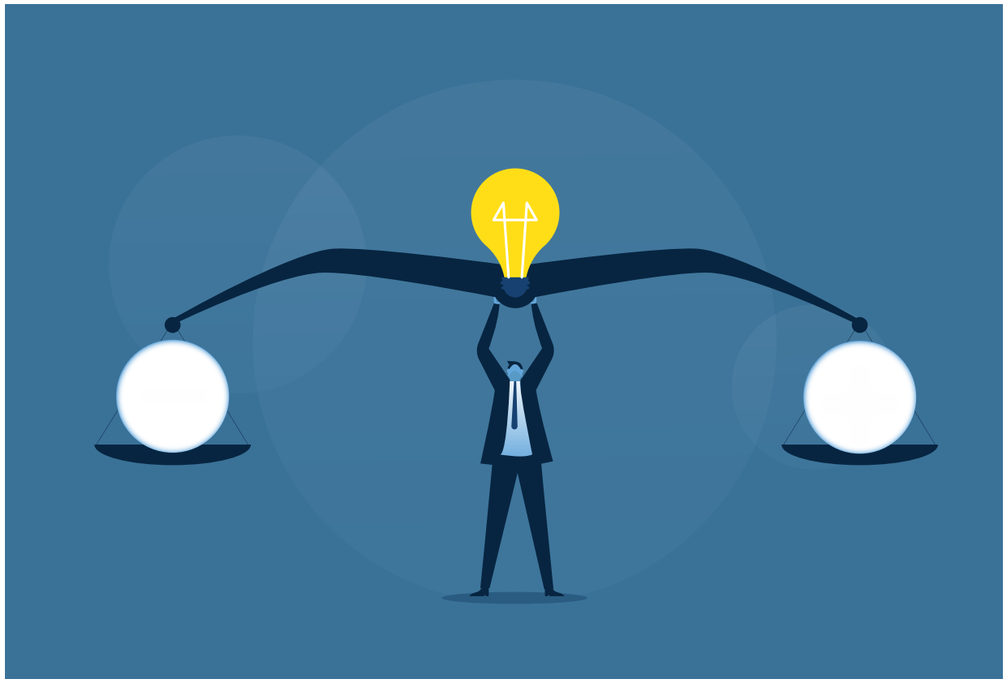
Use cases for anonymization depend on nature and context of the data, or its intended collection and utilisation purpose.[7]

> The adoption of anonymising techniques entails an endeavour to balance the utility of the dataset while mitigation of the risk of re-identification.[8]

Anonymization is used for personal data which is sought to be used beyond the original limited purposes of collection. Personal information can be detached and discarded from the dataset to overcome data minimization, purpose limitation, data sharing and RTBF (Right to be Forgotten) hiccups, although it is never entirely feasible to effectively remove all connections with personal information as the attributes are always available in the original dataset, or may be published later as part of another dataset, leaving room for linkability.[9] Generally, the degree of anonymization is inversely proportional to the magnitude of the use case.[10] Some types of anonymization affect the probability of singling out, and some data may not be suited to anonymization at all.[11] Entities must identify tractable datasets for anonymization at the earliest stage before active processing, as well as look beyond, for more suited data protection interventions.[12]

1.2 Balancing Risk Mitigation and Identifiability

Anonymization retains some inherent risk and must be considered alongside the multileveled concept of identifiability. Identifiability refers to the degree to which a natural person can be distinguished from other members of a group[13] by linking available data attributes to arrive at their identity.[14] Data contains direct identifiers (e.g.: name, unique social ID), indirect identifiers (e.g.: email address, phone number) and non-identifiers (non-personal data which appears anonymous on the face of it such as location data and browsing history). The presence of any of the first two uniquely distinguishes individuals, whereas the latter too may be linked to identifiers and provide sufficient context to identify individuals.[15] While it is not possible to quantify risks, recognizing possible shortfalls facilitates their eventual mitigation.[16] Effective anonymization converges to reduce the possibility of singling out, inferences and linkability.[17]



2

Comparing anonymization techniques

This chapter explores the interrelation between anonymization standards and anonymization techniques, depending upon technological advancement. It assesses the efficacy of the primitive techniques of data masking, redaction, pseudonymization and data suppression as opposed to newer techniques such as homomorphic encryption and differential privacy.

Anonymisation standards across various geographies have emerged from the cutting-edge techniques available at the time. As is the nature of technology however, most standards haven't been able to keep up with the rapid evolution in security and privacy practices and continue to use primitive or less reliable forms of anonymisation techniques. Over the past couple of decades, Privacy Enhancing Technologies (PETs) have changed the way anonymisation as an intervention is used and deployed. From primitive techniques of Data Masking, Redaction, Pseudonymisation and Data Suppression, to the present-day technologies of Homomorphic Encryption and Differential Privacy, technologists have redefined the standards of security best practices. Adopting innovative, safe and functional techniques based on organisational and regulatory requirements is the key objective of drafting useful standards. Before exploring how other geographies have enacted these standards based on various benchmarks and objectives, it is important to go over some of the most popular anonymisation tools currently in use.

Although techniques such as data masking and redaction, are no longer considered adequate tools for anonymising data, it had the set the stage for significant innovation in the domain of Privacy Enhancing Technologies at large. Through either obscuring, deleting, shuffling or adding pseudonyms in the place of true values, techniques such as a redaction, data suppression, swapping and pseudonymisation have enabled innovation in anonymisation standards, that has led to development of tools such as K-Anonymity and its later edit, L-diversity. K-Anonymity, for example, is a technique where the data is modified to ensure that each row of data is wholly indistinguishable from at least k-1 other rows in the dataset. This is done by generalising the data in each column in such a way, that each row of data looks indistinguishable from at least k-1 other rows of data. Here k can be any number, usually ranging from 3 to 20 or 21. The value of k decides the utility of data versus the degree of security and privacy. The higher the k value, the higher the

privacy. L-diversity on the other hand, is used to enhance the level of protection awarded to k-anonymous sets of data. The goal of L-diversity is to ensure that sensitive information within each k-anonymous group is diverse enough to protect against re-identification attacks. Tools such as t-closeness further extended the use case of L-diversity, by ensuring each k-anonymous group is close to the overall distribution across the entire data set, making each set more and more similar.

However, over the past half decade, techniques such as Differential Privacy and Homomorphic Encryption have now set the bar for what is considered adequate anonymisation interventions. Types of homomorphic encryption ranges from Partially Homomorphic Encryption (PHE) that supports one type of mathematical computation, Somewhat Homomorphic Encryption (SHE) that support a few types of mathematical computations, to Fully Homomorphic Encryption (FHE) that supports a host of complex computational challenges. Although it must be pointed out that FHE's that provide adequate performance is yet a problem waiting to be solved. Tools such a Differential Privacy (DP) is another that has been widely adopted and piloted across various industries. It works by introducing noise or "randomness" into the dataset, thereby reducing the ability to identify specific individuals. The objective is to ensure that the presence of a single data point does not have a say in the final computational outcome.

Homomorphic encryption enables one to perform analytics or computations on an encrypted dataset without having to decrypt it.

Newer interventions such as federated learning and synthetic data have further pushed the bar in what is considered optimal security and privacy best practice. Synthetic Data artificially generates highly nuanced data sets for analytics without the presence of any personally identifiable information in such sets. Even with some minor drawbacks, such interventions are setting the standard for future privacy practices.

3

Global perspective: Regulatory guidelines and best practices

This chapter explores best practices and standards for anonymization laid down by regulators in other countries, delving into the extent to which anonymization meets the requirements of Data Protection in that jurisdiction, and whether techniques beyond it in scope and effectiveness are preferred. The anonymization techniques discussed in the previous chapter may find mention here, depending upon the thresholds of prescribed by regulators. The following examples analyse the diverse approaches of these regulators to the standard of anonymization, governance of anonymization techniques, and risk-mitigation measures.

Singapore

Anonymization is covered by the Personal Data Protection Act (PDPA),[18] Guide to Basic Data Anonymisation Techniques,[19] Advisory Guidelines on the PDPA for Selected Topics[20] and the Guide to Basic Anonymization.[21] Notably, anonymised data falls outside the scope of Parts 3 to 6A of the PDPA as it is no longer considered personal data, although periodic reviews are required to assess the effectiveness of the anonymization. The scope of anonymised data is broader than the GDPR under the PDPA, referring to any dataset upon which anonymization techniques and safeguards[22] have been deployed to prevent any serious possibility of re-identification.[23] There is no legally differentiated standard prescribed for reversible and irreversible techniques, though their variants have been technically recognized.[24] Therefore, even as the Basic Guide acknowledges that anonymization may not be absolute and irreversible, the threshold of 'serious risk of re-identifiability' excludes reliable techniques able to pursue the objectives of anonymization better than itself, such as Differential Privacy and synthetic data.[25]

The PDPC sets out the legal requirements for effective anonymization in its Advisory Guidelines, and the technical recommendations for the steps involved through its Basic Guide. Organizations must remove direct identifiers, and alter or remove indirect identifiers in addition to the data recipient implementing additional safeguards towards the ends of access limitation, purpose limitation, retention policies, data minimisation and governance of the process. Where anonymised data is to be shared continuously over a period of time, periodic reviews should ideally be conducted to assess the adequacy of the anonymization technique and the robustness of risk management measures.[26] Organizations should strive to achieve a minimum k-anonymity value of 56 and of 3 for external and internal sharing respectively. Where this is not possible due to the need to preserve the granularity of data, they should put in place stringent safeguards such as tracking the lineage of data to mitigate re-identification risks. Certain risk mitigation controls in the form of access management, encryption, regular review of users, incident management, data sharing agreements, and internal governance controls should be

> PDPC recommends five technical steps to reach anonymised data: (1) knowing one's data, (2) de-identifying it, (3) applying anonymization techniques, (4) computing risks, and (5) and managing them.[27]

implemented for both the anonymised data and the identity mapping tables.

Since all datasets cannot be effectively or meaningfully anonymised, organizations are advised to consider the following while selecting an appropriate technique: (1) the nature and type of data, (2) potential impact on individuals, (3) nature of use and extent of disclosure, (4) public knowledge and personal knowledge, and (5) data recipient's ability and motivation to re-identify.[28] The effectiveness of anonymization is assessed against the motivated intruder test as a 'starting point', in addition to the possibilities of unintentional disclosure and data breach.[29] At the same time, newer privacy-preserving techniques are also able to meet the threshold of this test while providing more focused utility, simplifying privacy-by-design compliance and increasing organizational accountability.[30]

Canada

Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) does not refer to anonymization except as an alternative to destruction or erasure of personal information when it is no longer required.[31] Guidance indicates that while anonymised data falls outside the purview of PIPEDA, pseudonymised data may still be within its scope.[32] The incipient Bill C-27, presently under consideration,[33] seeks to remedy this situation by enacting the Consumer Privacy Protection Act (CPPA), Personal Information and Data Protection Tribunal Act (PIDTA), and Artificial Intelligence and Data Act (AIDA). The CPPA defines and distinguishes between the terms 'anonymize' and 'de-identify' in the following manner: 'anonymize' indicates a permanent and irreversible modification of personal information such that no individual can be identified from it by any means, whereas 'de-identify' refers to modification of personal information such that no individual can be directly identified from it despite a residual risk of re-identification.[34]

CPPA proposes to regulate whether anonymization is effective by referencing generally accepted best practices, and if answered affirmatively, it would be excluded from the scope of the Act.[35] On the other hand, de-identified data is regulated in its use, disclosure, and proportionality of technical and administrative measures used in the process.[36] Therefore, CPPA notably diverges from PIPEDA in positing the control of these forms of non-personal information. Though re-identification by organizations is strictly restricted in the case of de-identified data,[37] the regulator itself considers involuntary risks: information can never be truly de-identified, due to the simple reason that too much secondary data is available which, when combined with the de-identified data, can enable the identification of individuals.[38] When considering a user's reasonable expectation of privacy, entities must take into account what the entirety of the data taken together may reveal about the personal habits and choices of the individual behind the data.[39]

There is significant secondary legislation and literature available on de-identification. The Canadian government has released guidance in the form of Privacy Implementation Notices for deploying de-identification as a privacy preserving technique[40] and for reducing the risk of re-identification,[41] although applicable only to government institutions. They have been advised to (1) suppress, mask or redact direct identifiers, (2) assess the risk of re-identification through indirect identifiers, (3) determine the appropriate minimum cell size when releasing data in tables, (4) document the process, and (5) plan for a regular, ongoing and periodic assessment of re-identification risk.[42] Additionally, the Information and Privacy Commissioner of Ontario published the De-identification Guidelines for Structured Data in 2016 with a detailed nine-step process to balance utility and privacy while de-identifying a dataset.[43] The Canadian Anonymization Network (CANON) has also comprehensively laid down recommendations in its Report on Practices for Generating Non-identifiable Data (March 2021) for the following use cases: open release of non-identifiable data, release of non-identifiable data to an external party, custodian controlled access by third party to data, multi-party processing and internal data sharing.[44]

The proposed Act's definition of anonymization is quite onerous, with CANON describing it as setting "an extremely high and practically unworkable threshold."[45] This is especially in comparison to the jurisdictions like the GDPR, UK and Ireland with a 'reasonably foreseeable' standard for mitigating the risk of re-identification. The requirement to anonymize in accordance with 'generally accepted best practices' also takes away from organizations the flexibility to use novel and even stronger anonymization techniques simply because they do not fit within this description yet.[46]

European Union

The threshold of anonymization has been laid down by the Article 29 Working Party (WP29) Opinion and General Data Protection Regulation (GDPR) separately. In 2007, the Working Party adopted anonymization guidance[47] by interpreting Recital 26 of the EU Data Protection Directive, 1995.[48] It defines 'anonymous data' as any information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, taking account of all the 'means likely reasonably to be used' either by the controller or by any other person to identify that individual.[49] In 2014, WP29 added a stipulation that for data to be truly anonymised, data controllers must delete the original dataset.[50] The GDPR, however, has no such requirement.[51]

> ICO acknowledges the existence of a 'spectrum of identifiability' depending upon (1) the means reasonably likely to be used for anonymization, considering the data and its environment, context, scope and purposes of the processing, and technical and organisational measures applied, and (2) the identifiability risk considering motivation, competence needed, cost and time required, available technologies, and legal gateways and likelihood of their use.[56]

United Kingdom

The Information Commissioner's Office (ICO) has published a Code of Practice[52] and a series of PET guidance chapters concerning anonymization.[53] The Data Protection Act (DPA) does not directly refer to anonymization, but when read together with Recital 26 of the UK GDPR and the European Data Protection Directive (95/46/EC), indicates that data rendered anonymous in such a way that the data subject is no longer identifiable – whether by the data controller or by any other

person, considering all means likely reasonably to be used to identify that individual[54] – is not personal data and falls outside the scope of DPA.[55]

Organizations are also expected to have in place a comprehensive and effective governance structure, addressing staff training, responsibility for overseeing the process, mechanism to identify appropriate use cases for anonymization, knowledge management, joint efforts with similar organizations, privacy impact assessment (PIA), transparency, review, and disaster recovery.[57]

Ireland

The Data Protection Commission has published a Guidance Note on Anonymisation and Pseudonymisation[58] wherein effective anonymization is achieved when the data subject cannot be identified or identifiable taking into account all possible means that may be used for this purpose. The differentiating threshold between pseudonymization, which is still considered personal data, and anonymization, to which data protection principles would not apply, is whether the source data is deleted or retained at the time of applying the technique. Organizations are not expected to prove that it is impossible to identify a person from the anonymised data, but the identifiability test, involving direct or indirect identifiers as laid out by WP29 must be implemented with due regard to methods by which re-identification may take place.[59] It has been recommended that while anonymising data, the following be given due regard to - (1) sensitivity of the personal data, (2) who a motivated intruder may be, (3) how likely attempts at re-identification are, (4) other information or personal knowledge which the intruder may have access to enabling data linking, and (5) appropriateness of the anonymization technique considering the nature of the data.[60] On the sliding scale of low to high risk of identifiability, the DPC lists randomisation, permutation, generalisation, masking and pseudonymization respectively. It also encourages the use of these techniques in conjunction to reduce identifiability risks to the greatest possible extent.

The differentiating threshold between pseudonymization, which is still considered personal data, and anonymization, to which data protection principles would not apply, is whether the source data is deleted or retained at the time of applying the technique

France

The National Commission on Informatics and Liberty (CNIL) has adverted to WP29 guidance on anonymization techniques, like UK and Ireland[61] It lists some brief pointers for organizations to balance effective anonymisation with preservation of utility of data, and apply randomisation or generalisation accordingly. Anonymization would be achieved where individualisation, correlation and inference are not possible (WP29 standard)[62] Where there is a data breach post-publication of an anonymised dataset, the organization is required to remove it, inform CNIL and where risk to rights and freedoms of individuals is high, inform the people concerned[63]

Spain

The Spanish Data Protection Agency (AEPD) guidelines for anonymization techniques refer to six privacy-by-design principles to be kept in mind for anonymization processes: (1) proactive implementation as against reactive use, (2) privacy by default through planning of final granularity of anonymised data, (3) objective privacy considering the threshold of residual risk of re-identification, (4) full functionality to prevent distortion of data, (5) privacy preservation during information life cycle, and (6) training and information imparted to each actor in the anonymization process.[64] AEPD recommends differential privacy,[65] hashing algorithms and encryption mechanisms as more robust forms of anonymization, capable of achieving irreversibility,[66] but as a rule of thumb, requires organisations to use anonymization techniques together with data perturbation and data deletion, after checking the viability of the process through a pilot test.[67] Organizations are also expected to regularly evaluate the risk of re-identification as AEPD acknowledges that this processor is not entirely irreversible, and the risk associated with it can never be zero.[68]

Switzerland

While there is no regulator guidance on anonymization techniques, the Centre for Expertise in Social Sciences (FORS) has published an indicative guide of criteria organizations should endeavour to address while anonymising data. These include (1) nature and type of data, (2) resources required to anonymise the data, (3) sensitivity and potential harm caused to individuals if data is revealed, (4) future users of data, and (5) guarantees to be given to individuals about the use of their anonymised data. Depending upon the risk of harm to individuals, greater levels of anonymization, stronger promises regarding anonymization and access, and stricter controls must be established.[69]

Apart from these jurisdictions, it should be noted that the International Organization for Standardization (ISO) has published a new privacy-enhancing data deidentification framework in 2022 for identifying and mitigating re-identification risks and risks associated with the lifecycle of de-identified data.[70] National bodies are expected to adopt this standard through administrative and technical controls, contextual controls, data minimization and transformation, documented procedures and response mechanisms.[71] Essentially, this would enable the refinement of available anonymisation techniques from a globally agreed upon foundational launchpad.

4

Conclusion

The preceding sections have shed light on the framework available for anonymization from a technical and legal standpoint, as well as factors to be taken into account to balance security, usability and performance. Considering the best practices being followed in other jurisdictions and technological innovation, this chapter intends to make recommendations for appropriate standards which may be deployed by the Indian regulator or through industry self-regulation for the selection and implementation of anonymisation techniques.

Data anonymisation has come a long way from its original application to siloed data applications, to the present state of widespread sharing and combining of data.[72] As previously discussed, primitive techniques, such as pseudonymization and data masking, have proven to be the least impressive among anonymization operations due to their vulnerability to easy infiltration. However, they may still be used within organizations with highly secure communication lines for simple analytics or in very specific contexts.

On the other end of the spectrum, cutting edge techniques comprising of homomorphic encryption, differential privacy and synthetic data should be preferred by entities as they are able to approach dimensional data securely. Very few jurisdictions such as Spain and Singapore actively incorporate these techniques as the benchmark, and in this gap, India has the opportunity to emerge as the leader. While some newer interventions may be resource intensive, organizations should consider the goal of security worth the price. Significant data fiduciaries should ideally undertake the added labour of newer interventions due to the vast and intensive data processing carried out by them. Akin to Singapore's guidance on anonymization, legal and technical recommendations may be made for entities in terms of factors to be taken into account to select an anonymization technique, steps to mitigate risks, and graded threshold of anonymised data. Reference may also be made to Canadian guidelines on how to balance a user's reasonable expectation of privacy with utility of data, and ICO's motivated intruder test to define a threshold for effective anonymization. Canada's proposed CPPA also differentiates between the thresholds of anonymization applicable to and data protection measures to be afforded to de-identified data, anonymised data, and pseudonymised data, which may serve as a reference point for creating a risk-based framework. A graded approach as contained in the ISO deidentification framework may be created considering the amount of critical data and children's data, and like the AEPD's recommendations, organizations should be advised to regularly assess the risks to anonymised data and their data governance initiatives.

It is generally not considered prudent to create a prescriptive policy when it comes to anonymisation standards; rather an indicative set of guidelines may be published, with enough leg room for sectoral regulators to intervene with specific suggestions. Further, industry may be given the opportunity to self-regulate based on specific needs and requirements of the time.

REFERENCES

[1] GDPR, Recital 26, https://gdpr-info.eu/recitals/no-26/

[2] Alexandra Stam and Brian Kleiner, "Data anonymisation: legal, ethical, and strategic considerations", https://forscenter.ch/wp-content/uploads/2020/06/kleinerstam_fg11_anonymisation1_v1.0-1.pdf

[3] ibid p 5.

[4] Sharp Cookie Advisors, "Anonymization and GDPR compliance; an overview," https://www.gdprsummary.com/anonymization-and-gdpr/

[5] Lawrence Lessig, "Code Version 2.0," https://tigerprints.clemson.edu/cgi/viewcontent.cgi?article=1183&context=cheer

[6] ICO, "Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance" https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf

[7] ICO, "Introduction to Anonymization : Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance," HYPERLINK "https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf"https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf

[8] AEPD, "10 Misunderstandings related to anonymization," https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf

[9] ibid

[10] MeitY, "Draft Guidelines for Anonymization of Data," https://www.medianama.com/wp-content/uploads/2022/09/AoD_Guidelines_Draft_V1.0_July_22.pdf

[11] AEPD, "10 Misunderstandings related to anonymization," https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf

[12] ibid

[13] Article 29 Data Protection Working Party, "Opinion 4/2007 on the concept of personal data," https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[14] Data Protection Commission (DPC), "Guidance on Anonymisation and Pseudonymisation" https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf

[15] Personal Data Protection Commission (PDPC), "Guide to Basic Data Anonymization Techniques" https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf

[16] Data Protection Commission, 'Guidance on Anonymisation and Pseudonymisation' (June 2019) [https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf] page 6.

[17] ICO, ”Chapter 2: How do we ensure anonymisation is effective” HYPERLINK “https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf”https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

[18] Singapore- Personal Data Protection Act, https://sso.agc.gov.sg/Act/PDPA2012

[19] PDPC, ”Guide to Basic Data Anonymization Techniques” https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Other-Guides/Guide-to-Anonymisation_v1-(250118).pdf

[20] PDPC, ”Advisory Guidelines on the PDPA for Selected Topics,“ https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Selected-Topics/Advisory-Guidelines-on-the-PDPA-for-Selected-Topics-17-May-2022.pdf ch 3.

[21] PDPC, ”Guide to Basic Anonymization,” https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/Guide-to-Basic-Anonymisation-31-March-2022.ashx

[22] ibid

[23] PDPC, ”Advisory Guidelines on the PDPA for Selected Topics,“ https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Selected-Topics/Advisory-Guidelines-on-the-PDPA-for-Selected-Topics-17-May-2022.pdf 3.6

[24] Josh Lee, ”The value of differential privacy in establishing an intermediate legal standard for anonymisation in Singapore’s data protection landscape,” https://lawtech.asia/the-value-of-differential-privacy-in-establishing-an-intermediate-legal-standard-for-anonymisation-in-singapores-data-protection-landscape/

[25] ibid

[26] PDPC, ”Advisory Guidelines on the PDPA for Selected Topics,“ https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Selected-Topics/Advisory-Guidelines-on-the-PDPA-for-Selected-Topics-17-May-2022.pdf 3.7

[27] PDPC, ”Guide to Basic Anonymization,” https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/Guide-to-Basic-Anonymisation-31-March-2022.ashx

[28] PDPC, ”Advisory Guidelines on the PDPA for Selected Topics,“ https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Selected-Topics/Advisory-Guidelines-on-the-PDPA-for-Selected-Topics-17-May-2022.pdf

[29] PDPC, ”Advisory Guidelines on the PDPA for Selected Topics,“ https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Advisory-Guidelines/AG-on-Selected-Topics/Advisory-Guidelines-on-the-PDPA-for-Selected-Topics-17-May-2022.pdf 3.73.36 -3.40

[30] Josh Lee, ”The value of differential privacy in establishing an intermediate legal standard for anonymisation in Singapore’s data protection landscape,” https://lawtech.asia/the-value-of-differential-privacy-in-establishing-an-intermediate-legal-standard-for-anonymisation-in-singapores-data-protection-landscape/

[31] Canada - Personal Information Protection and Electronic Documents Act (PIPEDA), Sch.1- 4.5.3 https://laws-lois.justice.gc.ca/pdf/p-8.6.pdf

[32] Office of the Privacy Commissioner of Canada, ”A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act,“ https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/consent_201605/#fn77-rf

[33] Canada - Draft Digital Charter Implementation Act 2022, https://www.parl.ca/legisinfo/en/bill/44-1/c-27

[34] ibid s 2(1)

[35] ibid

[36] ibid s 74

[37] ibid s 75

[38] Office of the Privacy Commissioner of Canada, ”A discussion paper exploring potential enhancements to consent under the Personal Information Protection and Electronic Documents Act,“ https://www.priv.gc.ca/en/opc-actions-and-decisions/research/explore-privacy-research/2016/consent_201605/#fn77-rf; Paul Ohm, “Broken Promises of Privacy: Responding to the surprising failure of anonymization,” http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006

[39] Canada- R v Spencer (2016) SCC 43, https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/14233/index.do

[40] “Privacy Implementation Notice 2023-01: De-identification,” https://www.canada.ca/en/treasury-board-secretariat/services/access-information-privacy/access-information-privacy-notices/2023-01-de-identification.html

[41] https://www.canada.ca/en/treasury-board-secretariat/services/access-information-privacy/access-information-privacy-notices/2020-03-protecting-privacy-releasing-information-about-small-number-individuals.html

[42] ”Privacy Implementation Notice 2020-03: Protecting privacy when releasing information about a small number of individuals,” https://www.canada.ca/en/treasury-board-secretariat/services/access-information-privacy/access-information-privacy-notices/2020-03-protecting-privacy-releasing-information-about-small-number-individuals.html

[43] Information and Privacy Commissioner of Ontario, ” De-identification Guidelines for Structured Data,” HYPERLINK “https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf”https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf

[44] Canadian Anonymization Network, ” Practices for Generating Non-identifiable Data,” HYPERLINK “https://deidentify.ca/wp-content/uploads/2021/08/CANON-OPC-Project-Final-Report-v9.pdf”https://deidentify.ca/wp-content/uploads/2021/08/CANON-OPC-Project-Final-Report-v9.pdf

[45] Canadian Anonymization Network, ”Working Draft on Proposed amendments to de-identification and anonymization provisions in the Digital Charter Implementation Act, 2022 (Bill C-27),” HYPERLINK “https://deidentify.ca/wp-content/uploads/2022/09/Working-Draft-CANON-Proposed-Amendments-to-Bill-C-27-re-De-identify-and-Anonymize-Sept-30-2022.pdf”https://deidentify.ca/wp-content/uploads/2022/09/Working-Draft-CANON-Proposed-Amendments-to-Bill-C-27-re-De-identify-and-Anonymize-Sept-30-2022.pdf

[46] Barry B Sookman, ”CPPA: problems and criticisms – anonymization and pseudonymization of personal information,” https://www.mccarthy.ca/en/insights/blogs/techlex/cppa-problems-and-criticisms-anonymization-and-pseudonymization-personal-information

[47] European Union - Article 29 Data Protection Working Party, ”Opinion 4/2007 on the concept of personal data,” https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[48] European Council, ”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 on the free movement of such data,” HYPERLINK “https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML”https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML

[49] European Union - Article 29 Data Protection Working Party, ”Opinion 4/2007 on the concept of personal data,” https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf

[50] Article 29 Data Protection Working Party, ”Opinion 05/2014 on Anonymisation Techniques,“ https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

[51] GDPR, Recital 26, https://gdpr-info.eu/recitals/no-26/

[52] ICO, ” Anonymisation: managing data protection risk code of practice,” HYPERLINK “https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf”https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

[53] ICO, “Introduction to Anonymization : Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance,” https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf; ICO, ”Chapter 2: How do we ensure anonymisation is effective” https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

[54] United Kingdom - R (on the application of the Department of Health) v Information Commissioner (Admin) (2011) EWHC 1430 https://www.bailii.org/ew/cases/EWHC/Admin/2011/1430.html

[55] ICO, ”Anonymisation: managing data protection risk code of practice,” https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

[56] ICO, ”Chapter 2: How do we ensure anonymisation is effective” https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

[57] ICO, ” Anonymisation: managing data protection risk code of practice,” https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf p 41.

[58] Data Protection Commission, ”Guidance Note on Anonymisation and Pseudonymisation,” https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf

[59] Article 29 Data Protection Working Party, ”Opinion 4/2007 on the concept of personal data,“ https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf p 12.

[60] Data Protection Commission, ”Guidance Note on Anonymisation and Pseudonymisation,” https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf p 8-10.

[61] IAPP, ”A practical guide to anonymization standards across the EU and UK,” https://iapp.org/news/a/a-practical-guide-to-anonymization-standards-across-the-eu-and-uk/

[62] Article 29 Data Protection Working Party, ”Opinion 05/2014 on Anonymisation Techniques,“ https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

[63] CNIL, ” L’anonymisation de données personnelles,” https://www.cnil.fr/fr/lanonymisation-de-donnees-personnelles

[64] AEPD, ”Guidelines and Guarantees in the Anonymization Procedures of Personal Data,” HYPERLINK “https://www.aepd.es/documento/guia-orientaciones-procedimientos-anonimizacion.pdf”https://www.aepd.es/documento/guia-orientaciones-procedimientos-anonimizacion.pdf p 3, 4.

[65] AEPD, ”Anonymisation and pseudonymisation (II): Differential privacy,” https://www.aepd.es/en/prensa-y-comunicacion/blog/anonymisation-and-pseudonymisation-ii-differential-privacy

[66] AEPD, ”Guidelines and Guarantees in the Anonymization Procedures of Personal Data,” https://www.aepd.es/documento/guia-orientaciones-procedimientos-anonimizacion.pdf p 18, 19.

[67] ibid

[68] AEPD, “10 Misunderstandings related to anonymization,” https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf p 4, 5

[69] Alexandra Stam and Brian Kleiner, “Data anonymisation: legal, ethical, and strategic considerations”, https://forscenter.ch/wp-content/uploads/2020/06/kleinerstam_fg11_anonymisation1_v1.0-1.pdf p 9, 10.

[70] ISO, ” Information security, cybersecurity and privacy protection – Privacy enhancing data de-identification framework,” https://www.iso.org/standard/71677.html

[71] IAPP, ”A New Standard for Anonymization,” https://iapp.org/news/a/a-new-standard-for-anonymization/

[72] Shoshanna Solomon, ”Data is up for grabs under outdated Israeli privacy law, think tank says,” https://www.timesofisrael.com/data-is-up-for-grabs-under-outdated-israeli-privacy-law-think-tank-says/

A nasscom Initiative

Data Security Council of India (DSCI) is a premier industry body on data protection in India, setup by nasscom, committed to making the cyberspace safe, secure and trusted by establishing best practices, standards and initiatives in cybersecurity and privacy. DSCI brings together governments and their agencies, industry sectors including ITBPM, BFSI, telecom, industry associations, data protection authorities and think-tanks for policy advocacy, thought leadership, capacity building and outreach initiatives. For more info, please visit www.dsci.in

DATA SECURITY COUNCIL OF INDIA

Nasscom Campus, Fourth Floor, Plot. No. 7-10, Sector 126, Noida, UP - 201303

+91-120-4990253 | research@dsci.in | www.dsci.in

DSCI_Connect dsci.connect dsci.connect data-security-council-of-india dscivideo