International Journal of Multidisciplinary Research and Development
www.allsubjectjournal.com
ISSN Online: 2349-4182 ISSN Print: 2349-5979
Received: 30-09-2023, Accepted: 15-10-2023, Published: 31-10-2023
Volume 10, Issue 10, 2023, Page No. 85-89

Anonymisation and data privacy

Walubita Luwabelwa
Department of Compliance, Chief Compliance Officer, Stanbic Bank Zambia Limited, Zambia

Abstract
The relationship between anonymity and privacy ought on the face of it to be compliemntary. In reality, it has not always been clear cut; for the most part it has been tenous, and at worst egregious. Society has witnessed a proliferation of technological invasions in people's day to day relations and communications, raising concerns around the extent of personal exposure of one's information to data collectors, data collectors and even law enforcers. The attempt at drawing a good balance between the two concepts has exercised minds of social engineers, lawyers, and policy makers alike. Unsurprisngly, the response has both been technological and legislative in the form of anonymity. Anonymity thereby allows for access whilst drawing a line in the sand against datication and commodification of personal information in the information market. Ultimately, anonymity allows of the enjoyment of civil liberties without untoward. Left unchecked, anonymity has been blamed for promoting nefarious actions such as Dark Web and flips the coin onto its face for cyber security. This article steers away from the philosophical arguments of data privacy, nor the civil liberties discourse but focusses more on the positive promise of anonymisation when used positively. It interrogates whether the downsides of anonymity merits the call for less deidentification techniques like pseudonomisation.

Keywords: Anonymisation, anonymity, privacy

Introduction

Businesses and organizations today hold large amounts of data used for different purposes, primarily for better customer service and business efficiencies. The value of these large volumes of personal data have made businesses and organisations alike attractive sources for attackers and othe unauthorised third parties. Data by this token is ubiquitous propelled by the realisation of its intrinsic value (scientific, societal, economic).

Egragious abuses into individual's privacy led to regulatory action to balace interests of large "big data" companies, consumer rights activists, regulators, politicians and legal ractitioners to resolve matters of what data privacy means vis-à-vis rights of data subjects. Ultimatley data controllers, data processors and regulators alike aim to find a way to maintain the data's utility while adequately reducing the risk of sensitive information being leaked.

Cosequently, global standards were promulgated that saw the introduction of a cost for the use and collection of data, driving the need to have techniques that would transform this high value asset into a form that will not need adherence to strict data privacy legislation and cost, while maintaining its use with minimal distortion; the rise of anonymisation. This is because data protection law does not apply to truly anonymous information, inanimate entities nor to the deceased.

Personal data does not always have to be anonymised as there now exists international principles for legitimate processing of personal data. Anonymity is defined as the state of being not identifiable within a set of subjects, and its attainment leads to hughest levels of privacy as it enables an individual's ability to keep and protect their identity.

Privacy in its most common parlance relates to the an individual's right to control access to their information; confidentiality relates to protecting an individual's personally identifiable data and anonymity is refraining from collecting an indivdiual's personal idetifiers, direct or indirect.

The three concepts though related are distinct in meaning and implications for a data processor. This article focusses on the aspects of anonymity, its differences with data privacy and pseudonomysation, common misconcetions and also consides the probative value and risks of not implementing proper anonymization techniques.

Anonymisation

The increase in pervasive tracking, monitoring, and recording has fundamentally changed our day to day lives. Anonymity, has become a more accepted recourse in dealing with this growing trend, albeit becoming more and more difficult to attain due to social, economic, and political drivers. This growing social control has earned terms such as "informationalisation"; natural causative result is a yearning for increased privacy protection *vis-a-vis* information privacy and data protection laws.

Anonymity is generally understood on a spectrum of controlled visibility from one end of identifiability to the other end of unknowability. It is often synomised with concepts of untraceability, or unreachability. Some commentarors have described anonymity as a tool of resistance against visibility and trackability, without compromising participatory communication as it allows for some level of presence that may be seen but not datafied.

Due the stricter regimes surrounding processing of personal data, however, the quest for anonymisation has garnered more attention for the obvious reason that it no longer becomes subject to data protection requirements. Namely, the benefits of anonymisation include increased span in use of the data, reduced cost of compliance to data protection requirements, easier disclosure and reduced proper use limitations.

The question as to whether information has been trully anonymised more often than not is determined by surrounding facts relating to the means reasonably likely to be used to identify an individual that the information relates to. This will be covered later in this article.

Effective anonymisation is possible by using techniques that reduce the risks of identifying individuals to a sufficiently remote level that they effectively anonymise the information. Exceptional instances may also render it almost impossible to anonymise data, thereby still necessitating the support of confidentiality requirements.

The principle of Proactive Responsibility requires the data controller to undertake a study of the inherent risk of re-identification of the people who the data refer to and implement measures to manage it. The aim of that analysis is to tailor the anonymisation process to the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons. In other words, to achieve an adequate balance between the need to obtain results with a certain degree of fidelity and the potential cost of processing for citizens' rights and freedoms.

The end goal of anonymising information is to move personal data into a form of information that is outside the scope of data protection legislation.

1. Defining Anonymous Information

Data protection generally does not expressly define or categorise anonymous information. This understandably is motivated by the fact that it is often a question of fact; who is using the data and for what purpose, and that anonymous information is an end result of anonymisation.

Notwithstanding, anonymous information has been defined as data which does not relate to an identified or identifiable individual (i.e. data that is not personal data). It is 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.

Relatedly, anonymisation is the process of turning personal data into anonymous information so that an individual is not identifiable.

The Zambian Data Privacy Act defines anonymisation as the:

> *"process of removing direct and indirect personal identifiers that may lead to an individual being identified.*

The operative word in the Zambian legislation is process, and hence gives it a wide understanding to the reader that the method and tools utilised to remove information outside the ambit of data protection laws is what is most critical.

A similar view has been taken by the UK GDPR law which does not consicely define anonymous information. The provision states that

> *'…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.'*

Two important data privacy principles to consider at this point identifiability are the reasonably likely test.

2. Identifiability

Underpinning the concept of anonymisation is the concept of identifiability. Data may include bothe direct and indirect identifiers. A direct identifier refers to specific information that references to an individual (name, identification number etc), while an indirect identifier is any piece of information (geographical position, opinions etc) that could be used, either individually or in combination with other indirect identifiers, by someone that has knowledge about that individual with the purpose of re-identifying an individual in the dataset.

Attributes of records according to the type of information they contain can be classified in three generic groups:

a. Key identifiers

Fields which unequivocally identify the data subjects (name, ID, passport number, telephone number, etc.). Those types of data must be removed from the anonymised records.

b. Quasi-identifiers

Fields which, in themselves and in isolation, do not identify an individual but which, if grouped together with other quasi-identifiers, could unequivocally identify a subject. Anonymisation techniques work on those data, removing fields which are not necessary for processing (in application of the principle of minimisation), aggregating them or generalising them.

c. Sensitive attributes

Fields which contain data which could have a greater impact on the privacy of a specific individual, including the special categories of data, and which must not be linked to the data subject they belong to (illnesses, medical treatments, income level, etc.). That information may be of great interest in the object of the data processing, but unless there is some legitimation for it, it must be dissociated from a specific subject.

A common technique for achieving anonymity is K-anonymity. The concept is premised on the idea that individuals' data is pooled in a larger group; information in the group could correspond to any single member, thus masking the identity of the individual or individuals in question. Thus it is a measure of the risk that external agents can obtain information of a personal nature from anonymised data.

It is not always possible to anonymise data in all circumstances; lowering the re-identification risk below a previously defined threshold whilst retaining a useful dataset for specific processing. Similarly, anonymous data today may not be anonymous data in future with changes in available information and processes.

It is however important to caution against equating anonymisation with automation: whilst automated tools can be used during the anonymisation process, there is a role for human expert intervention.

Data minimisation as a principle requires that the legitimate purpose be clearly understood by the data processor/controller; if that purpose can be fulfilled by using personal data as opposed to anonymous data and the acceptability of the reidentification risk.

3. The 'reasonably likely' test

The re-identification likelihood is the probability in a given dataset of re-identifying an individual by turning anonymised data back into personal data through the use of data matching or similar techniques. This risk is never completely eliminated; the records in any dataset have a probability of being re-identified based on how possible it is to single them out, which risk can be assessed by using an effective anonymisation process.

Article 29 Working Party provided some guidance on the re-identification risk on whether is within acceptable limits:

3.1 Confirming that the anonymized data does not manifest any of the three properties

a. **Singling out:** some records of an individual in the dataset can be isolated**:**
b. **Linkability:** at least two records concerning the same data subject or a group of data subjects can be linked
c. **Inference:** one or more attribute values can be deduced with significant probability

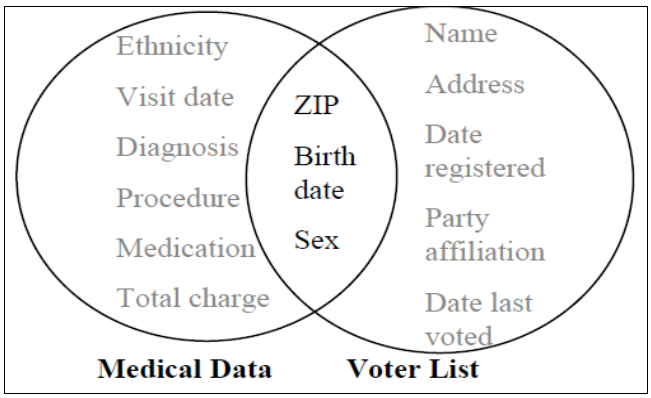
3.2 Performing a re-identification risk analysis.

What then is the position of encryption?

Encryption is often misunderstood as producing anonymous data, which is not often correct. Using the distinction above, the possibility of re-identificaion must be effectively removed to pass the threshold for anonymisation. Short of that, it will be at best pseudonymous data.

The encryption process uses secret keys to transform the information in a way that reduces the risk of misuse, while keeping confidentiality for a given period of time. Some commentators have noted that a major weakness of encryption is that it has a mirror process of decryption; thus the secret keys used for decryption is additional information which can make the personal data readable and consequently identifiable.

Different factors may affect the likelihood of re-identification, such as using a large sample size. Sweeney demonstrates re-identification by directly linking shared attributes (ZIP code, birth date and gender to medical information) linking diagnosis, procedures, and medications to particularly named individuals. He gives an example of a record of William Weld, who was governor of Massachusetts and lived in Cambridges Massachusetts. According to the Cambridge Voter list, six people had his birth date; three of them were men, and he was the only one in his 5 digit ZIP code as below:



4. Benefits of Anonymisation

The attraction towards anonymisation have already been highlighted earlier in this article. The ultimate benefit is the maximisation in the usage of data as it is no longer subject to the strict requirements of data privacy legislation relating to personal data.

Some have cited specific benefits of anonymisation that can help processors, controllers and other users in the data processing chain. Effective anonymisation can, among other things, therefore help one to:

- better understand the legal requirements about information you hold and intend to share or disclose;
- improve decision-making and risk reduction and management processes;
- adopt a 'data protection by design' approach;
- reduce reputational risks arising from inappropriate or insecure disclosure or publication of personal data;
- develop greater confidence in publishing anonymous information and navigating potentially challenging FOI requests involving personal data;
- achieve greater transparency, public trust and confidence as a result of proper publication of anonymous information without abrogating privacy protection;
- incentivise researchers and others to use anonymous information instead of personal data, wherever this is possible;
- realise economic and societal benefits deriving from the availability of rich data sources.

In light Techniques and approaches that are designed to turn personal data into anonymous information constitute processing operations performed on that data. For example, when you create aggregate statistical information from…

5. Processing and Anonymisation

Anonymisation has been defined earlier in this article. It was earlier stressed that anonymisation emphasis on the process or manner in which the conversion of information is done with the ultimate result of not falling within data protection legislation. Under data protection, this term is 'processing'.

The criticality of this term is that it defines the boundaries that delimitate who can carry out this process, where and how this conversion can occur.

Anonymisation in effect lends itself to techniques and approaches that turn personal data into anonymous information, constituting processing operations: aggregation of statistical information as an adaptation or alteration. The presumption is that the purpose for the processing satisfies the data protection principles and adequately documented.

English Data Protection Act defines processing as "…an operation or set of operations performed on information or on sets of information, such as –

Collection, recording, structuring or storage

a. adaptation or alteration
b. retrieval, consultation or use
c. disclosure by transmission, dissemination or otherwise making available,
d. alignment or combination, or
e. restriction, erasure or destruction.

Under the Zambian regime, processing means an operation or a set of operations which is or are performed on personal data, whether or not by automatic means, including the collection, recording or holding of the data or the carrying out of any operation or set of operations on data, including—

a. organisation, adaptation or alteration of the data;
b. retrieval, consultation or use of the data;
c. alignment, combination, blocking, erasure or destruction of the data; or
d. disclosure of the information or data by transmission, dissemination or otherwise making available;

The two definitions are materially similar in the scope of what amounts to processing. The notable difference is the United Kingdom adds 'storage' and 'structuring' while the Zambian legislation extends to 'holding' and 'blocking'

Anonymisation as the Dark Horse

Attacks against anonymisation can take several forms, and it is emperative that data prcessors and controllers alike are aware of which form this may take:

1. deliberate attempts at re-identification;
2. unintended attempts at re-identification;
3. data breaches; or
4. releasing data to the public.

Pseudonymisation

In its simplest form, pseudonymisation is a method that replaces or removes information that identifies an individual. The aim is to avoid re-identification of an individual through the additional information or other forms of identifiers, which places a high burden on a processor to ensure the additional information is stored separately with the appropriate controls.

The UK GDPR defines the term as

> *'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'*

This is to ensure that it is not possible to re-identify an individual from use of the separately held additional information, or indeed any other information.

Pseudonymisation should be thought of as a security and risk mitigation measure, not as an anonymisation technique by itself; a type of processing designed to reduce data protection risk, but not eliminate it.

Why then distinguish between anonymisation and pseudonymisation?

The importance of the distinction is that anonymous information is not personal data, while pseudonymous data is personal data.

The use of 'additional information' can lead to the identification of the individuals, under pseudonomysation, which is why pseudonymous personal data is still personal data. In comparison, anonymous data cannot be associated to specific individuals and once data is truly anonymised, individuals are no longer identifiable and the data will not fall within the scope of personal data legislation.

Similarly, the nature of offences falling under each; the offence of re-identifatication under the UK DPA relates to pseudonymised data rather than anonymised data. It is important to note that not all juridictions with data protection legislation have an express re-identification offence; Zambia for instance only contains the offence of general unlawful disclosure of sensitive personal data to another person.

The distinction is aptly put in the following terms

Anonymisation means that individuals are not identifiable and cannot be re-identified by any means reasonably likely to be used (ie, the risk of re-identification is sufficiently remote). Anonymous information is not personal data and data protection law does not apply. Pseudonymisation means that individuals are not identifiable from the dataset itself, but can be identified by referring to other information held separately. Pseudonymous data is therefore still personal data and data protection law applies.

The danger facing processors alike is whether data held has crossed the threshold of anonymisation. Where the data is re-identifiable, depending on the de-identification and storage techniques of the re-identification data sets, it is often safer to presume it is pseudonomyised rather than anonymised data.

1. Benefits of Pseudonymisation

Pseudonymisation has recognised benefits. With pseudonomysied information, a data processor can perform a greater number of general analysis tasks not easily available to personal data. It aids in the process of data protection by design principles, is an appropriate security measure, personal data breach notification tool and safety measure.

Conclusion

Information by its nature wants to ebb freely between sender and recipient. In the information economy, data continues to be a valuable currency with benefits derived from identification that facilitates direct marketing, data sovereignty and implementation of cyber security measures. Driven by the incessant desire to package information into quantifiable formats for tabulation and analysis, what has became obvious is the need to develop a regime of privacy protection for data considered worthy of protection.

This article has shown that a data controller must ensure the privacy of the subjects whose data is being processed is achieved. It is not always possible to guarantee anonymity as some common fields present in in more focussed data elements may raise the risk of re-identification. What is paramount is to take a risk-based approach of balancing out the freedoms of data subjects against the use of appropriate anonymisation techniques which meet the requirement for accountability.

Reference

1. Adrian Yanes, Privacy and Anonymity, Aalto University, 2014.
2. Aggravating factors of reasonable likely test of re-identification includes strength of the encryption algorithm/key, information leaks, implementation issues, quantum of encrypted data, technologies such as quantum computing. European Data Protection Supervisor, AEPD-EDPS, Joint Paper on 10 Misunderstandings Related to Anonymisation, 2021.

3. Alan F. Westlin: Privacy and Freedom; Atheneum
4. AEPD, Anonymity as a Privacy Measure, Available at http://www.agpd.es; https://sedeagpd.gob.es
5. Barth-Jones D. The 're-identification' of Governor William Weld's medical information: a critical re-examination of health data identification risks and privacy protections, then and now. Then and Now, 2012.s
6. Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance, Information Commissioner's Office, 2021.
7. European Data Protection Supervisor, AEPD-EDPS, Joint Paper on 10 Misunderstandings Related to Anonymisation
8. General Data Protection Regulation (EU) 2016/679
9. Helen Nissenbaum, "The Meaning of Anonymity in an Information Age," The Information Society, 1999, 15(2).
10. http://www.agpd.es; https://sedeagpd.gob.es
11. https://www.immuta.com/blog/k-anonymity-everything-you-need-to- know-2021-guide/
12. Information technology Security Techniques Evaluation Criteria for IT security, 2005.
13. Introduction to anonymisation Draft anonymisation, pseudonymisation and privacy enhancing technologies guidance, Information Commissioner's Office, 2021.
14. Khaled El Emam, Luk Arbuckle. Anonymizing Health Data
15. Sweeney L, K-anonymity. a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002, 10(5).
16. Melis, Rachel. "Anonymity Versus Privacy in a Control Society," in "Information/Control: Control in the Age of Post-Truth," eds. Stacy E. Wood, James Lowry, and Andrew J Lau. Special issue, Journal of Critical Library and Information Studies, 2019, 2(2). DOI:10.24242/jclis.v2i2.75.
17. Michael A. Froomkin, "Anonymity in the Balance," in Digital Anonymity and the Law: Tensions and Dimensions, ed. C. Nicoll, J.E.J. Prins, and M.J.M. van Dellen (The Hague: T.M.C. Asser Press, 2003.
18. Pfitzman A, amp Kohntopp M. Anonymity, Unobservability and pseudonymity-a proposal for terminology. In H. Federrath, (Ed.), 16 Proceedings of the International Workshop on Design Issues in Anonymity and Unobservabilit, 2001.
19. Preserving Privacy in Artificial Intelligence Applications through Anonymisation of Sensitive Data, Deloitte
20. Somolinos R, Cristóbal A. Muñoz Carrero, M.E. Hernando Pérez, M. Pascual Carrasco, R. Sánchez de Madariaga, O. Moreno Gil, J.A. Fragua Méndez, F. López Rodríguez, C. H. Salvador. Pseudonimización de información clínica para uso secundario. Aplicación en un caso práctico ISO/EN 13606, Unidad de Investigación en Telemedicina y e-Salud, Instituto Carlos III, Madrid, 2014.