BUSINESS INTELLIGENCE

The role of anonymisation and pseudonymisation in data privacy

Mastek

Table of contents

Introduction

Privacy. It's a hot topic and it's going to get hotter. Breaches and hacks are hitting headlines; a new and more stringent EU data protection regulation is coming soon (GDPR); Safe Harbor has been overturned; consumers are behaving in revealing ways (20% are using ad blockers and it's growing fast); and countless surveys are telling us what we do with our customers' data is critical to whether they trust and will therefore buy or consume, our products and services. Organisations across industries that include finance, retail, healthcare, energy, manufacturing, and other sectors have long recognized the advantages of having access to consumer data. But with the exponential growth of data, privacy law changes and the advent of complex technologies that are making it easier to collect and process consumer information, data privacy has become a critical commercial issue.

With the aim of providing citizens added security against identity theft and misuse of their personal information, governments across the globe have established data protection regulations that outline the standards that should be strictly implemented in the course of collecting, storing, processing, and transferring data, and particularly when dealing with personally identifiable information (PII).

The need to comply with these stringent regulations coupled with the risk of having to shoulder hefty fines for non-compliance is forcing business enterprises and public agencies to invest in IT solutions that are designed to enhance the security and privacy of the data they hold while still allowing them to process such data as they see fit within the bounds of existing regulatory directives.

Data anonymisation and pseudonymisation have come to the forefront as viable technological measures that can significantly ease concerns on data privacy and protection.

Anonymisation and pseudonymisation explained

Generally Anonymisation is defined as the process of manipulating data such that the resulting information is stripped of any elements that will identify the data subjects. Once anonymisation techniques are applied to raw data, the remaining data set should be devoid of attributes that:

- single out a specific individual;
- link to other sensitive information about the subjects included in the data; or
- allow the data user to deduce a subject's identity.

Effective anonymisation usually entails omitting or deleting the pertinent information that could establish the identity of an individual.

Pseudonymisation is a related de-identification process that also purposely conceals the identity of individual data subjects. It does so by replacing unique identifiers (typically this might be the most common personal information such as name, date and place of birth, social security or NHS number, and so on) in a dataset with artificial attributes using security techniques as blurring, tokenisation, short codes and masking. Inasmuch as the mapping pseudonyms of the subject identities are retained by the data controller (or the organisation that controls the original data), pseudonymisation is generally considered a reversible process, wherein retracing of data back to its original or identifiable state is possible.

In practice the end use case of the data and organisation's security policies define the level of acceptable re-identification risk. Thus more complex blends of both anonymisation and pseudonymisation techniques exist in real-world data anonymisation transformations.

It's worth noting that the terms anonymisation and pseudonymisation are not fixed definitions and slight variations are normal depending on the industry context within which the technique is used, or the regulatory standard to which it is held to. Most data protection laws and standards do not provide specific technical guidelines on how full anonymisation can be achieved nor what level of pseudonymisation supports acceptable levels of re-identification risk.

Why anonymise or pseudonymise data?

With the sheer amount of data that can be mined today, it's no understatement to say that data is the new gold. Others opt to call it the new currency or the new oil. Whatever name is given to it though, one thing is certain: data has become a valuable business asset that many organisations are only now learning to use to boost efficiency and drive revenue.

But with the issues of consumer trust and privacy continuously hounding organisations at every turn, data anonymisation (and in that we are including, pseudonymisation) technology has become a must-have for any enterprise's security portfolio. Yes, implementation of data anonymisation requires internal resources and technological tools. But the investment is well worth it when you consider the benefits.

Anonymisation and pseudonymisation allow organisations to:

- **Achieve compliance with data privacy regulations.** Anonymisation and pseudonymisation raise an organisation's level of adherence to data privacy laws. In the General Data Protection Regulation (GDPR) the maximum fine for non-compliance will be 4% of the errant company's global gross revenue or €20 million, whichever is greater. These are hefty potential fines and change the cost-benefit analysis significantly when considering implementing privacy-enhancing technologies.

- **Mitigate risk of security breaches.** Keeping data secure should be an organisation's top priority. But when a data breach does happen, focus is placed on what type of data has fallen into the wrong hands. When properly de-identified, anonymised, or pseudonymised data is transferred or exposed to non-authorized personnel, PII or sensitive information is not compromised. This reduces the risk of disclosure and misuse of personal data by malicious insiders or cyber criminals, and therefore spares the organisation from (or substantially reduces) the added financial costs and reputation damage of breach notification, legal liabilities, customer defection, loss of opportunity, and others.

- **Maximise value of data.** Data can extensively benefit both individuals (e.g. health care, improvement of retail services) and organisations (e.g. profiling, marketing, forecasting). But these applications also call for data to be processed and shared, which puts the privacy and protection of individuals' personal data at risk. Anonymisation and pseudonymisation techniques give organisations the capability to utilise data while still adhering to privacy regulations.

- **Build consumer trust.** Even more than the governments, consumers too are anxious about how their privacy rights are being upheld. And with no end in sight for security breaches, they have every reason to be so. Customers understand that data is vital to enable businesses to offer personalised experience and anticipate their needs, but organisations should be transparent about the safeguards they have put in place for proper data management. Implementing data pseudonymisation or anonymisation would go a long way into giving that much needed assurance to customers that their data is kept private. Once organisations recognize the numerous advantages of data anonymisation and pseudonymisation, the question that remains is not 'Why anonymise data?' but 'Why NOT anonymise data?

Anonymisation and pseudonymisation in light of evolving EU regulations

Data anonymisation has attracted greater interest in the recent shakeup of regulations regarding EU data transfers from EU member countries to the USA and then the release of the draft GDPR.

Schrems

Entities across Europe and US suffered a huge setback in October 2015 when the Court of Justice of the European Union CJEU deemed the Safe Harbor Decision as invalid in its landmark ruling on the Schrems v. (Irish) Data Protection Commissioner Case. Though we now have Privacy Shield, it's focused people's attention on data transfers and left people unsure whether Privacy Shield would withstand another challenge from Schrems or anyone else. So firms are looking at alternatives for both EU-US transfers for all transfers of data outside the EU. When Schrems was overturned leading UK-based law firm Pinsent Masons LLP in its client note 'Safe Harbour – What Now?' suggested organisations who continue to rely on Safe Harbor could find themselves in breach of contracts that require protection of personal data and in violation of privacy regulations. Along with encrypting and tokenising, Pinsent Masons further recommended anonymising personal data as a "technical safeguard" that DPAs can consider as capable of providing "adequate safeguards for certain transfers." Organisations should considering these techniques for data transfers out of the EU as a means of reducing and even replacing their reliance on the untested protections of the new Privacy Shield.

GDPR

GDPR (General Data Protection Regulation) will be enforced on 25th May 2018. The regulation will severely alter ways in which companies collect, store, process and protect the personal information of customer.

GDPR creates enough incentives for enterprises to psuedonymise data:

- Article 32 states that organisations should implement appropriate technical measures to ensure a level of security by applying techniques like pseudonymisation and encryption of personal data. Enterprises implementing these techniques will have an advantage when it comes to demonstrating compliance to the regulation.

- Article 25 of the regulation advocates the principle of data protection by design and default. It encourages building privacy into products and solutions right from the design phase rather than tackling privacy later on. The article also mentions that Pseudonymisation can be one of the key features for implementing this key principle of data protection by design and default.

- No enterprise wants a data breach but sometimes it is a harsh reality. Under GDPR the data controller is under an obligation to notify the Supervisory authority of any data breach within 72 hours and if the risk is high then controller is also under an obligation to notify the data subject. Pseudonymisation and encryption reduce this risk by a considerable extent by storing personally identifiable data separately protected by appropriate safeguards and making the data very difficult to de-identify.

- Enterprises implementing pseudonymisation will be at an advantage to protect data and demonstrate compliance to the regulation.

Anonymisation in the context of test data masking

The information age has brought with it a plethora of data from across a vast spectrum of sources –business apps, data sensors, customer databases, smart devices, and social media, among others. An enterprise's ability to harness such information in the development of future knowledge determines that organisation's capacity to stay competitive and relevant in the industry and its potential to increase revenue and boost profit.

But with data security issues to seriously consider, it's crucial to find the right balance between maintaining consumer privacy and retaining utility of data. Anonymisation and/or pseudonymisation effectively resolves the impasse between the objectives of data processing and the importance of privacy protection by providing a valuable layer of security for sensitive, private and onfidential information found in test and development databases. The process of replacing personal information in test databases using anonymisation or pseudonymisation techniques is generally known as data masking. An effective masking software can successfully de-sensitise databases byreplacing existing sensitive information in a test database with other values that may appear real but are actually artificial. Thus it can generate a similarly structured and consistent data set to the original.

Outsiders and unauthorized insiders, who might somehow gain access to the data with the goal of using or misusing it, will instead find it unusable. But intended users can still glean information from the remaining data and will find it functional for testing, development, training, QA, research, presentation of statistical data, and other similar applications in non-production environments.

For example, pharmaceutical and healthcare companies submit confidential patient data for medical research to evaluate the efficacy of new treatments or asses clinical trial results. Working with cloaked databases where key attributes (such as name, NI number) are removed, and quasi-identifiers (such as birthday, post code) are pseudonymised does not prevent researchers from finding relevant results in their study.

In financial services, another example, masked data is used for testing new transaction systems or fraud modelling. Maintaining the table structures, speed of masking and consistently masking is the key to effective test data creation, especially in Big Data landscape.

But these are just two examples and for industries with financial and health data, classed as highly sensitive. Taking such techniques and using them on all personal data is a logical next step for privacy and data security across many sectors. And use cases for masking are even more varied because anonymisation/pseudonymisation software and data masking tools are not only suitable for personal data (PII and PHI) but also for many types confidential business information.

Anonymisation in the context of transferring to cloud services and 3rd parties

Cloud computing has many recognized benefits: reduced upfront hardware costs, on-demand scalable computing capacity, increased business agility, disaster recovery capabilities, and more. But with a number of stringent data privacy and security regulations to contend with and the fear of sending data outside your own perimeter, not a lot of organisations are prepared to fully embrace cloud adoption.

Enterprises are understandably hesitant to store and share information in the cloud because this immediately raises the question of privacy and confidentiality.

The same privacy and security concerns are brought up when data is transferred to 3rd parties, such as when companies outsource software application development and testing to offshore locations, give data to marketing agencies or other suppliers.

Anonymisation is a powerful tool that can be used to preserve confidentiality and reduce privacy risk while still allowing data to be stored, utilised, and analysed in a public cloud. Organisations can process anonymised data in the cloud or transfer pseudonymised data to 3rd parties with far less concern that the data may be inadvertently exposed to or captured by unauthorised users. Once data is anonymised or de-sensitised for identifying information, its value and use to others is reduced dramatically.

Achieving privacy goals with anonymisation

As the business landscape continues to move towards a data-rich environment, the immense challenge for organisations is to make the most use of the available information while maintaining compliance to the ever-tightening government regulations on data privacy and protection. That's certainly a tall order for any business enterprise. But not if you have the right technological tools to back you up – anonymisation and pseudonymisation.

Data anonymisation and pseudonymisation have been identified as powerful processes that can pave the way for achieving privacy goals. But not just any pseudonymisation technology will do. It is imperative for organisations to opt for anonymisation or pseudonymisation software that adopts the right combination of techniques and algorithms to effectively increase data security and privacy for their data, analytics projects and information systems.

Our solution: kameleonTM

Kameleon™ is an Enterprise Pseudonymisation and Anonymisation toolkit that seamlessly plugs into a variety of data flow architectures, de-sensitising the fields passing through it that are marked as sensitive while leaving the other fields as is. The KameleonTM software application de-sensitises the data for analysis, data sharing and other instances where "distinguishable" but not "identifiable" data can be used to protect the privacy of individuals.

From our ISO-compliant methods, strong encryption algorithms, service-oriented architecture, and easy implementation model, we've come up with a privacy-by-design solution that can help organisations comply with data privacy regulations without negative impact to existing databases and applications.

Learn more about KameleonTM and find out how it can fit into your business' data security and privacy needs. Contact us at **info@mastek.com.**

About **Mastek**

Mastek is an enterprise digital transformation specialist that engineers excellence for customers in the UK, US and India. We enable large-scale business change programmes through our service offerings, which include application development, support and testing, BI and analytics, agile consulting and digital commerce.

Whether it is creating new applications, modernising existing ones or recovering failing projects, we help enterprises to navigate the digital landscape and stay competitive.

Learn more by visiting **www.mastek.com**

Mastek

Global Headquarters
Mastek UK Ltd
Pennant House
2 Napier Court
Reading, RG1 8BW

+44 (0)118 903 5700
www.mastek.com

IndigoBlue

3–4a Little Portland Street, London, W1W 7JB

+44 (0)20 7692 4832
www.indigoblue.co.uk

TAISTech

15601 Dallas Pkwy
Suite 250
Dallas, TX 75001

+1 972 521 3063
www.taistech.com

Mastek Ltd

#106, SDF IV, Seepz
Andheri (East)
Mumbai – 400 096

+91 22 6695 2222
www.mastek.com/in

Follow us on: