Preserving Privacy in Artificial Intelligence Applications through Anonymization of Sensitive Data

Executive Summary

Anonymization can play a critical role in building trust in Artificial Intelligence (AI) and its applications within society. Various anonymization techniques have been used in order to shield individual privacy in the context of these datasets. Traditional approaches to anonymization have focused on "data masking" (obfuscation of content, i.e., data is encrypted and difficult to understand without help) for generating test data. In contexts where data processors ensure that personal data are sufficiently anonymized, they may convince more data subjects to agree on having their data collected, or on more information, than would be the case without anonymization. Such data could be of particularly high quality, significantly improving the performance of Artificial Intelligence systems. AI is powered by data, so access to a great amount of it – compliant with privacy regulations – is key to unlock the full potential of AI. The robustness, and therefore usefulness, of AI systems is dependent on having access to high quality data.

Any organization serious about incorporating AI effectively into its processes, products and services should view anonymization as a crucial component of its data management strategy and analytics best practices. It is a critical primary step for using data securely in secondary processes. ❯

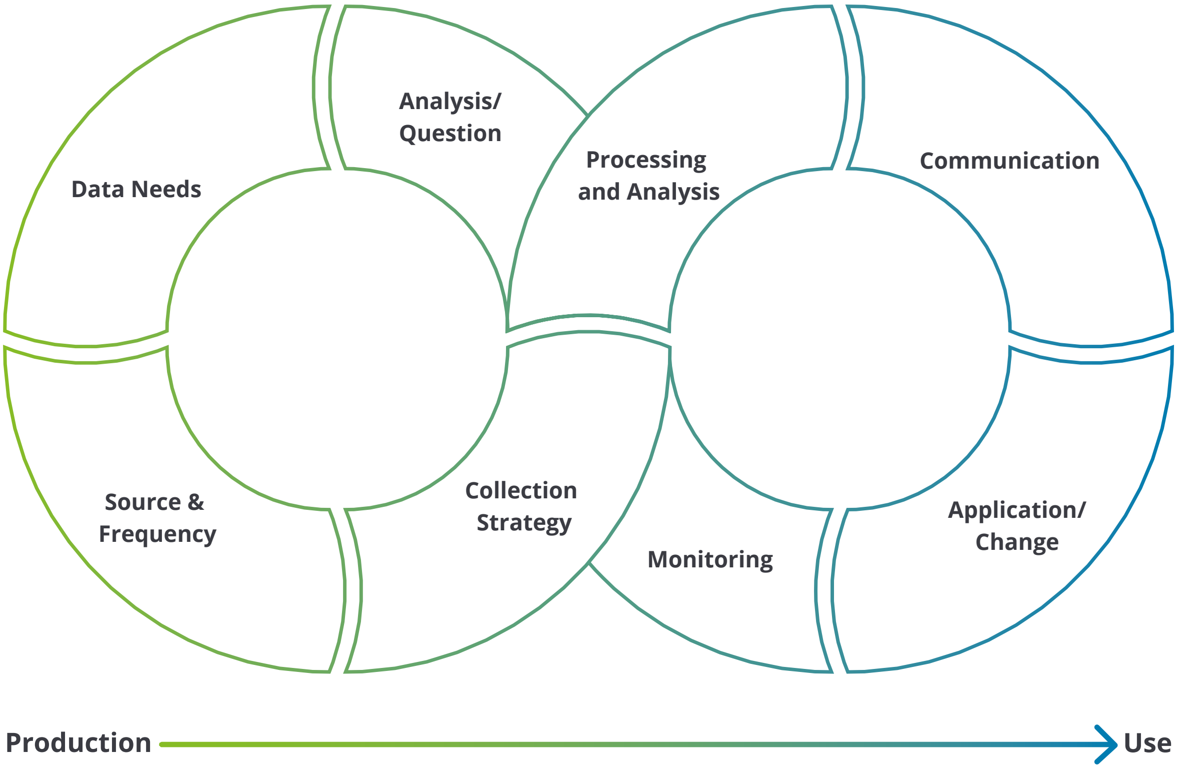
MAKING AN IMPACT THAT MATTERS since 1845

Data Makes the Modern World Go Around

The exchange of data is the currency of our time. Organizations have become increasingly skilled at monetizing data – and keen to collect more. The free flow of information has created many business opportunities – as well as opportunities for theft. Embarrassing data breaches and costly cyber-attacks give cause to re-think how to add value with data while still maintaining privacy.

The data value chain (see fig. 1) depicts the process of data creation and use from collection through re-use. Responsibly passing data along this value chain in its original form requires strict controls and data-sharing agreements. In many cases, anonymized data may fully meet the needs for insights, thus reducing the risk of accidental or malicious re-identification that expose personal information.

Fig. 1 – Data Value Chain



Varying Degrees of Obscurity

Anonymization is the process of manipulating data such that the resulting information is stripped of any elements that could identify the data subjects. Once anonymization techniques have been applied to sensitive data, it should no longer be possible to:

- Single out a specific individual
- Link to other sensitive information about the subjects included in the data
- Allow the data user to deduce a subject's identity

Effective anonymization usually includes techniques that omit or delete information that can be used to identify a specific individual.

A related de-identification process called pseudonymization also purposely conceals the identity of an individual subject, yet retains the ability to re-identify the dataset if later required. This is achieved by substituting unique identifiers with an artificial attribute using techniques such as tokenization and masking. Pseudonymization is a reversible process, whereas with anonymization, the individual identities are permanently lost.

It is worth noting that the terms anonymization and pseudonymization are fluid definitions and slight variations are normal depending on the industry context within which a technique is used, or the regulatory standard to which it is held. Most data protection laws and standards do not provide specific technical guidance on what constitutes adequate anonymization or what degree of anonymization acceptably limits re-identification risk.

Tab. 1 – Anonymized vs Pseudonymized Data

Anonymized Data	Pseudonymized Data
Identifying individuals is impossible or requires inordinate effort.	Individuals can be identified following the addition of information that is stored separately or publicly accessible.

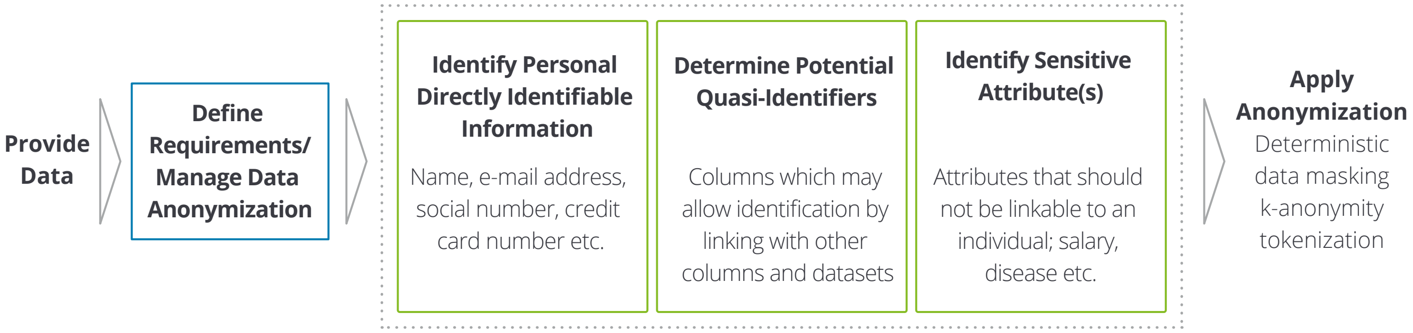
Different Obligations under GDPR

Data strategies for pseudo- or full anonymization are subject to different requirements under the General Data Protection Regulation (GDPR). Full anonymization effectively excuses the data collector/analyst from data stewardship obligations, while pseudonymization paints a mixed picture:

Tab. 2 – GDPR Obligation

GDPR Obligation	Anonymized	Pseudonymized
Provide notice to data subject	Not required	Required
Give right to be forgotten	Not required	Depends on degree of anonymization
Data retention limitations	Not required	Required
Data security	Not required	Partially met
Data breach notification	Not required	Depends on degree of anonymization
Documentation/record keeping obligation	Not required	Required
Obtain consent or have another legal basis	Not required	Not required but helpful

Fig. 2 – Data Anonymization Process


Provide Data
Define Requirements/ Manage Data Anonymization
Identify Personal Directly Identifiable Information
Name, e-mail address, social number, credit card number etc.
Determine Potential Quasi-Identifiers
Columns which may allow identification by linking with other columns and datasets
Identify Sensitive Attribute(s)
Attributes that should not be linkable to an individual; salary, disease etc.
Apply Anonymization
Deterministic data masking
k-anonymity
tokenization

Risk-Based Anonymization Process

After receiving the data from source, the attributes of the data are classified into direct identifiers (e.g. name, address, telephone number, license plate number, e-mail address) and quasi-identifiers (e.g. sex, date of birth or age, postal codes, ethnic origin, total income, profession, event dates, number of children, high level of diagnoses and procedures). After both direct identifiers and quasi-identifiers are identified, the sensitive attribute is defined, that must not be linked to the individual.

We then measure the risk of re-identification and relate this to a risk threshold representing the degree of risk we deem acceptable. A threshold value of zero would allow no useful data to be shared – whereas a value one means that no data has been anonymized. This means that the acceptable risk of re-identification is going to be some value larger than zero. The risk threshold would typically be higher when data is shared externally than within an institution, or when the data concerned relates to categories of personal data subject to stronger regulation such as health records.

The actual risk of re-identification is measured using techniques discussed later in this paper, compared to the threshold, then a decision can be made as to the acceptability of the risk for re-identification. It is worth mentioning that the risk of re-identification is dependent on the context itself, i.e. the actual re-identification risk is a function of the context and the data. The context represents the security, privacy, and contractual controls that are given; for example, one context can be a public data release whereas another example could be data that is provided to a researcher used within a secure environment.

Extensive documentation of the anonymization process (as depicted in fig. 2) ensures that questions about the assumptions can be answered later and to allow replicability of the analysis.

Techniques, and Technical and Organizational Requirements

Multiple anonymization techniques exist with varying degrees of robustness.

The Article 29 Working Party[1], an independent body advising the European Commission, has provided recent guidance in data anonymization techniques in which it recommends two approaches to confirm re-identification risk is within acceptable limits:

1. Either confirming that the anonymized data do not manifest any of the three following 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 (in the same or in two different databases)
   c. Inference: one or more attribute values can be deduced with significant probability

2. Or performing a re-identification risk analysis

However, a competing analysis of the Working Party's opinion argues that meeting all of the criteria listed under 1. above would result in datasets with limited utility – anonymized, but no longer useful. The natural conclusion is to perform the re-identification risk analysis. Furthermore, risk-based anonymization methods are consistent with recommendations from other authorities, such as the Information Commissioner's Office in the UK[2] or the Determination Methods under the HIPAA Privacy Rule in the US[3].

There may be situations where merely satisfying two of the three criteria will suffice for the intended task. For instance, when the intention is to anonymize three different datasets independently linked to one another, it may be favourable to neglect the risk of linkability to preserve referential integrity. Leveraging pseudonymization, it is possible to use an allocation table that links every plain-text item of data to one or more pseudonyms. The allocation table is, in effect, a key/token. Only those who have access to the key can link a pseudonym to the associated plain-text item of data by scanning the relevant entries. Similarly, pseudonymization can employ cryptographic scrambling techniques, which transform a plain-text item of data into one or more pseudonyms. The cryptographic technique is another form of key, through which re-identification can be restricted and managed. Restricted access to keys such as the allocation table or cryptographic scrambler is paramount.

A common misconception around pseudonymization is that removing or replacing attributes will result in anonymized data.

k-Anonymity

k-Anonymity obscures the individual identity of a data subject by grouping at least k individuals together. It is achieved by generalizing the attribute values to an extent such that each individual shares the same value. For example, by lowering the specificity of a location from a street to a postal code (or a city to a country) the location information is no longer distinct, applying to a higher number of data subjects. Similarly, specific birth dates could be generalized to birth months or years. Continuous numerical values (e.g., salaries, weight, height, or the dose of a medicine) could be "bucketized" into intervals (e.g., a salary range €20,000–€30,000). These methods are useful to avoid the distinct attribute values becoming quasi-identifiers.

With the same attributes shared by k users, it should be no longer possible to single out an individual within a group of k users (see the right section of table 3, where the attributes Postal Code, Age and Nationality were obscured, bucketized and suppressed respectively). However, k-Anonymity cannot ensure referential integrity as it eliminates the ability to link to other sources. It remains possible to link records by groups of k users, where the probability that two records of a group correspond to the same pseudo-identifier is 1/k.

Tab. 3 – Example of k-Anonymity for k=4

Original Data					k-Anonymized (k=4)				
	Quasi-Identifiers*			Sensitive Attribute		Quasi-Identifiers			Sensitive Attribute
ID	**Postal Code**	**Age**	**Nationality**	**Disease**	**ID**	**Postal Code**	**Age**	**Nationality**	**Disease**
1	13053	28	British	AIDS	1	130**	< 30	*	AIDS
2	13068	29	German	AIDS	2	130**	< 30	*	AIDS
3	13068	21	American	Tuberculosis	3	130**	< 30	*	Tuberculosis
4	13053	23	German	Tuberculosis	4	130**	< 30	*	Tuberculosis
5	14853	50	Indian	Diabetes	5	1485*	≥ 40	*	Diabetes
6	14853	55	British	AIDS	6	1485*	≥ 40	*	AIDS
7	14853	47	German	Tuberculosis	7	1485*	≥ 40	*	Tuberculosis
8	14853	49	German	Tuberculosis	8	130**	≥ 40	*	Tuberculosis
9	13053	31	German	Diabetes	9	130**	3*	*	Diabetes
10	13053	37	French	Diabetes	10	130**	3*	*	Diabetes
11	13068	36	American	Diabetes	11	130**	3*	*	Diabetes
12	13068	35	German	Diabetes	12	130**	3*	*	Diabetes

* A quasi-identifier is a data attribute that does not identify the individual on its own, but can do so in combination with other attributes of the dataset.

A common mistake is to artificially augment the value k by reducing the considered set of quasi-identifiers as seen in table 4.

A quasi-identifier is an attribute that, in combination with other attributes (also quasi-identifiers) can enable re-identification. Here, the attribute Nationality has not been considered as a quasi-identifier when setting the value of k. Thus, in this example we have k=1, as for example a single Indian age 40 and older from postal code 1485* can be singled out. Reducing quasi-identifiers makes it easier to build clusters of k-users due to the inherent power of identification associated to the other attributes (especially if some of them are sensitive or possess a very high entropy, as in the case of very rare attributes). If some attributes can be used to single out an individual in a cluster of k, then the generalization fails to protect some individuals.

Grouping a set of individuals with an uneven distribution of attributes is problematic in most cases. The impact of an individual's record on a dataset will vary: some will represent a significant fraction for the entries while the contributions of others remain fairly insignificant. It is therefore important to ensure a sufficiently high k parameter to avoid individuals representing a dominant portion of the entries in a cluster.

To eliminate risk further the anonymized data can be subject to ensuring stricter criteria such as L-diversity, which consider the number or distribution of sensitive attributes within a cluster, i.e. L-Diversity measures the diversity of sensitive values for each cluster in which they occur. It aims at having at least L different values for a sensitive attribute. In the previous example L-Diversity would be 1 as all individuals in their 30s living in postal code 130** have Diabetes.

In case an attacker knew an individual in the dataset with these quasi-identifiers they could identify their sensitive attribute.

Tab. 4 – Example of k-Anonymity for k=4 without considering Nationality

k-Anonymized (k=4)				
	Quasi-Identifiers			Sensitive Attribute
ID	**Postal Code**	**Age**	**Nationality**	**Disease**
1	130**	< 30	British	AIDS
2	130**	< 30	German	AIDS
3	130**	< 30	American	Tuberculosis
4	130**	< 30	German	Tuberculosis
5	1485*	≥ 40	Indian	Diabetes
6	1485*	≥ 40	British	AIDS
7	1485*	≥ 40	German	Tuberculosis
8	1485*	≥ 40	German	Tuberculosis
9	130**	3*	German	Diabetes
10	130**	3*	French	Diabetes
11	130**	3*	American	Diabetes
12	130**	3*	German	Diabetes

Pseudonymization

Pseudonymization replaces one attribute (typically a unique attribute, often the name of an individual) in a record with another value. The natural person is therefore still likely to be indirectly identifiable; as such, pseudonymization alone will not yield a truly anonymous dataset. Nevertheless, with many misconceptions and mistakes surrounding its use, it is an important topic to address.

Pseudonymization reduces the linkability of a dataset with the original identity of a data subject. In this sense, it is a useful security measure, if not an anonymization method. The result of pseudonymization can be either independent of the initial value or it can be derived from the original attribute values (e.g., a hash function or encryption scheme). The most popular pseudonymization techniques are:

- Encryption
- Hash function
- Deterministic encryption
- Tokenization

Pseudonymization still allows individual records to be singled out, as the individual is still identified by a (surrogate) unique attribute. This presents a clear advantage in terms of preserving linkability between records, linked on the pseudonymized instead of the original attribute. In fact, even if different pseudonymized attributes are used for the same data subject, linkability may still be possible by means of other attributes. The link is only definitively lost if no other attribute in the dataset can be used to identify the data subject and if every link between the original attribute and the pseudonymized attribute has been eliminated (including by deletion of the original data). In such cases, there can be no obvious cross-reference between datasets using different pseudonymized attributes.

The same linkability advantage presents a security vulnerability: inference attacks on the real identity of a data subject are possible within the dataset or across different databases that use the same pseudonymized attribute for an individual, or if pseudonyms are self-explanatory and do not mask the original identity of the data subject properly.

The most serious problem with pseudonymization is the misconception that removing or replacing one or more attributes will result in anonymized data. This has been disproven on many occasions, where original identity has been successfully derived from quasi-identifiers remaining in the dataset as just altering the private individual information does not prevent someone from identifying a data subject if values of other attributes are still capable of identifying an individual. In many cases it can be as easy to identify an individual in a pseudonymized dataset as with the original data. For instance, a significant study in 1997 by Latanya Sweeney[4], the co-author of the concept of k-anonymity, examined anonymized medical records that the US state of Massachusetts had released with all direct identifiers removed. Sweeney was able to re-identify most records using only the date of birth, sex, and postal code. She was, most notably, able to uniquely identify the Governor of Massachusetts and all his medical history. She additionally identified that, at that time, 87 percent of the US population could be identified by date of birth, sex and postal code.

Again, the appropriate anonymization approach will depend on the context. Nevertheless, the consequences may not always be clear at the outset. It is important to avoid using the same key in different databases to be able to reduce linkability if the data include easily linkable attributes that are also present with the other sources. The following example will illustrate the problem in more detail:

Two datasets of medical records contain the social security number, the gender, birth date, the postal code and additional medical data, such as the disease. Both datasets contain the social security number, the gender and the postal code. Imagine all attributes are anonymized in the same way in both datasets. If someone knows the date of birth, gender and postal code of an individual, it is feasible to link this de-identified medical record uniquely to the individual, and is able to learn sensitive medical data about this natural person.

Anonymization When Transferring to Cloud or Other 3rd Parties

While cloud computing promises many benefits – from reduced hardware costs to on-demand scalable capacity – it nonetheless involves transferring data beyond the firewall to 3rd party providers, which is an inherent security risk. Despite security efforts of cloud providers, some organizations are not prepared to fully embrace the lure of cloud. In the face of stringent privacy and security regulations, they are understandably reluctant to surrender their data to a provider outside the safe perimeter of the company's internal firewall.

Anonymization offers a powerful solution to address these concerns, preserving confidentiality and reducing privacy risk to transmission, storage and exchange of data in a public cloud. It offers organizations significant protection against inadvertent exposure of data to or capture by unauthorised users.

Re-Identification of Anonymized Data

Up to now we have focused on techniques to anonymize, or "de-identify" data, a useful assurance if deciding to benefit from cloud capabilities. Yet, what if we want to re-identify the original data subjects after having processed a dataset in the cloud? Re-identification or "de-anonymization" is the process of matching a previously anonymized dataset with other datasets (be they public, private, or maliciously obtained) to deduce the natural person to whom that data belongs. Re-identification can be a useful tool – akin to decryption – but it can also pose a problem to organizations, especially if their privacy policies hold them responsible to maintaining anonymity of data they have chosen to share. These are factors an organization must consider early in the decision process in determining how to and how much to de-identify both direct- and quasi-identifiers. The organization must also consider whether their pseudonymization algorithms (the lock) may have been developed such that they could be reverse-engineered (the key).

Risks of Re-Identification

The UK Office for National Statistics (ONS) conducts motivated intruder testing. They define success of this testing as "a small number of correct claims [re-identifications] with low confidence" rather than completely removing all risk that might make the data set unuseful. What would be the impact of failure to consider the anonymization principles, such as k-anonymity or sufficiently abundant pseudo-identifiers?

In a now well documented case identified by Anthony Tockar of Northwestern University in 2014, the New York City Taxi and Limousine Commission was issued a Freedom of Information Law (FOIL) request for information pertaining to all 2013 taxi rides[6]. That data included "pickup and drop-off times, locations, fare and tip amounts, as well as anonymized (hashed) versions of the taxi's license and medallion numbers." As discussed in the findings, a web search on "celebrities in taxis in Manhattan in 2013" was enough to find a picture, connect to quasi-identifiers in the data, and identify two specific celebrities who had been in the taxis, where they started and stopped, and how much they paid and tipped. All major breaches of privacy.

A final example shows how insufficient anonymization can open the door to malicious re-identification. As part of a research project into re-identification risks in 2017, Svea Eckert from The Guardian and Andreas Dewes, a data scientist, set up a fictional marketing company, reached out to market research companies and, with a little effort and no monetary renumeration, obtained a 30-day anonymized browsing history with 2 billion URLs of 3 million German users across 9 million different websites[7]. Combining this data with publicly available sources, such as social media handles, they were able to re-identify specific individuals, their sexual preferences (from visited sites), and other very personal information. The results of their experiment were presented at the DEFCON hacking conference in Las Vegas as a cautionary tale for others.

Besides the loss in reputation, companies found to be in violation of existing data protection regulation may face significant fines. Following the EU GDPR, violators may be fined up to 4 percent of global annual revenue of the preceding year in the case of serious infringements against data privacy.

Purposeful vs Malicious Re-Identification

Identity theft aside, there are many legitimate reasons for re-identification, such as identifying an original credit card number to refund a transaction, or researchers seeking insights through more specific information. Situations such as these could be resolved through the original data owner re-identifying the dataset, querying additional information, and anonymizing again before sharing the results with the requestor. Using the anonymization and tokenization techniques to create strong pseudo-identifiers described previously, a data owner could create a dataset to share with limited re-identification risks.

Unfortunately, with the steady flow of data breaches globally, the ability of a bad actor to succeed with malicious re-identification of a dataset is ever-present. In a European Union (EU) study released in January 2020, it was reported that there were 160 thousand data breaches reported under the GDPR in the 8-month period from May 2018 to Jan 2019[8]. Studies such as these underscore the importance of validating and quantifying the re-identification risk in a dataset.

Quantitative Validation Processes

Each anonymization method discussed in this paper utilizes statistical methods to calculate the probability of re-identification of the dataset. We have shown that proper identification and anonymization of quasi-identifiers is a subjective practice. A more quantitative way to predict the probability of re-identification is needed to measure the efficacy of the chosen anonymization technique, how it compares to industry norms for the type of data, and what kind of risk they are approving prior to dataset release. Several countries have already introduced measures to minimize the risk of re-identification. For instance, the Singapore Personal Data Protection Commission defines sensitive information and suggests techniques to anonymize them accordingly. Further, it defines the risk of re-identification in a probabilistic way.

To determine the probability of re-identification, the intended recipient (external recipients such as academics or internal recipients such as team members) for which the anonymization is obtained is a crucial factor to determine the necessary degree of re-identification probability so that data is shielded from unintended recipients, whether they are within or outside the organization.

The most dominant factor is the privacy model (such as k-anonymity), a combination of a concept to minimize re-identification risk and the corresponding metrics for measuring its success. The privacy model must take the relevant attacker models (e.g., prosecutor, Journalist or marketer models[9]) into consideration, which differ in how many individuals are targeted and whether the attacker knows about the data target's existince in the dataset or not. The privacy model is only valid for one table at a time – measuring the risk of re-identification for several tables at a time will leave some residual risk.

Despite the advantages of a quantitative approach, depending on the methodology used, in some circumstances it is not advisable to apply a metric, for instance, when generating synthetic data. Still, if there is some underlying structure to the data, a possible inference from the data should be considered for the risk assessment.

The intended recipients, the privacy model and applied methodologies must be jointly considered to calculate the final probability score. Re-identification risk will ultimately depend on the data and the specific use case. A guidance for a threshold for the overall metric is a probability of less than

0.09, which is suggested by the EMA[10]. If the probability is slightly higher than 0.09 the usage should be considered carefully – whereas if the probability is much higher, then the anonymization process should be repeated.

Motivated Intruder Testing

In conducting motivated intruder testing, the UK Office for National Statistics (ONS) makes use of "friendly intruders" to observe whether they can successfully recover the identity of anyone in the anonymized dataset. Testers can be internal resources or trusted partners who will attempt to re-identify a record set as a final phase of the anonymization process. "These intruders should have some background knowledge of the data similar to that of a typical user. However, they do not need to be specialist hackers with the capability of employing advanced data exploration techniques."[11]

Motivated intruder testing is an important tool in testing anonymization effectiveness prior to release outside the organization, as well as for internal datasets that need be anonymized prior to analysis. This assumes having utilized a quantitative validation process prior to testing, to achieve a level of certainty concerning the quality of anonymization. Rigorous procedures such as motivated intruder testing grow in importance with increasing sensitivity of the data, such as medical records. Organizational policies should clearly layout when just a quantitative analysis is required, or when a motivated intruder testing is warranted.

Conclusions

Data is ubiquitous – driven on the one side by sensors and gadgets, but much more by the realization what value can be derived from it – scientific, societal, economic. Plentiful and creative uses of data began to intrude on the privacy of individuals, with some egregious abuses triggering regulatory action. The debate ensued between "big data" companies, consumer protection groups, legal professionals, regulators and politicians about what data privacy means, the rights of the individual and which protections need to be put in place. As a result, various regulatory regimes have proposed and enacted restrictions and penalties into law, such as GDPR in Europe. Strict regulation and controls have introduced a new "cost" to collection and use of data, quickly transforming troves of data from an asset to a liability. This has given rise to interest in techniques to "de-risk" or "de-sensitize" data.

There are more motivations to do so than regulatory compliance alone. Proper anonymization techniques can instill trust from an organization to its customers and employees. Several techniques are available – anonymization/pseudonymization/synthetic data – each with their advantages and disadvantages, each suitable for different situations. Yet naïve implementation can be as potentially damaging as harbouring personal data to begin with.

For example, performing anonymization without considering re-identification exposes an organization to unintentional risk that could damage the organization's reputation. When determining anonymization methods, one size does not fit all. The specificities of data sets, the degree of required confidentiality, the risks of exposure – each of these will dictate the appropriate technique.

Deloitte is committed to ensuring the use of technology is trustworthy and ethical – for ourselves and our clients. Data privacy is a core competency, and partnering with best-in-class application developers along with a significant investment into a proprietary anonymization framework enables Deloitte and it's clients to achieve their analytical or test data needs while protecting individual privacy. We offer our clients a fundamental analysis of their data and the associated use cases. Then we design and implement a data anonymization process using various techniques of anonymization in line with the appropriate privacy models to meet the requirements of data protection laws. Furthermore, the possibility of inference on the anonymized dataset is significantly reduced but the data remains valuable for the specified use case.

Glossary

Terminology	Description
Anonymization	The conversion of personal data into anonymized data by applying a range of anonymization techniques.
Anonymized data	The resultant dataset after anonymization techniques have been applied in combination with the relevant risk assessments.
Attribute	Also referred to as data field, data column or variable. An information that can be found across the data records in a dataset.
Dataset	A set of data records. Conceptually similar to a table in a typical relational database.
Direct identifier	A data attribute that on its own identified an individual or has been specifically assigned to an individual.
Quasi-identifier	Columns which may allow identification by linking with other columns and datasets.

Sources

[1] https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf
[2] https://ico.org.uk/for-organisations/guide-to-data-protection/
[3] https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html
[4] https://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_06-2005_testimony_sweeney.pdf
[5] https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guidanceonintrudertesting
[6] https://perma.cc/5LZG-YZM8
[7] https://www.theguardian.com/technology/2017/aug/01/data-browsing-habits-brokers
[8] https://www.statista.com/chart/20566/personal-data-breaches-notified-per-eea-jurisdiction/
[9] https://www.lexjansen.com/phuse/2017/dh/DH09.pdf
[10] https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf
[11] https://www.ons.gov.uk/methodology/methodologytopicsandstatisticalconcepts/disclosurecontrol/guidanceonintrudertesting

Contacts

David Thogmartin
Director
Risk Advisory (Germany)
aiStudio | AI & Data Analytics | AI Institute
dthogmartin@deloitte.de

Richard Leaton
Senior Manager
Enterprise Data Architecture Leader (US)
Global and Strategic Services
rleaton@deloitte.com

Patrick Spitzer
Senior Consultant
Risk Advisory (Germany)
AI & Data Analytics
pspitzer@deloitte.de

Alexandros Melemenidis
Senior Consultant
Risk Advisory (Germany)
AI & Data Analytics
amelemenidis@deloitte.de

www.deloitte.com/de/aistudio

Deloitte.

This communication contains general information only, and none of Deloitte GmbH Wirtschaftsprüfungsgesellschaft or Deloitte Touche Tohmatsu Limited ("DTTL"), its global network of member firms or their related entities (collectively, the "Deloitte organization") is, by means of this communication, rendering professional advice or services. Before making any decision or taking any action that may affect your finances or your business, you should consult a qualified professional adviser.

No representations, warranties or undertakings (express or implied) are given as to the accuracy or completeness of the information in this communication, and none of DTTL, its member firms, related entities, employees or agents shall be liable or responsible for any loss or damage whatsoever arising directly or indirectly in connection with any person relying on this communication. DTTL and each of its member firms, and their related entities, are legally separate and independent entities.

Issue 02/2022