Draft 1

Privacy-Preserving Data Sharing: A Comprehensive Literature Review

20-06-2024

Summary:

This report provides an overview of the key concepts and techniques in privacy-preserving data analysis, emphasizing the importance of differential privacy, data anonymization, and data perturbation. Differential privacy, introduced by Dwork et al. in 2006, remains a cornerstone, ensuring that the inclusion or exclusion of any individual's data has minimal impact on the outcome of data analyses, thereby offering strong privacy guarantees. The privacy-preserving principles of autonomy, justice, non-maleficence, beneficence, and explicability are critical for maintaining ethical standards in data handling.

Furthermore, the report discusses the ethical foundations of privacy, stressing the need for continuous ethical reflexivity amidst rapid technological advancements. Compliance with international standards like ISO/IEC ensures that privacy-preserving methods meet global security benchmarks. The proposed frameworks integrate these techniques and principles to offer a robust approach to protecting individual privacy in data analysis, addressing the complex challenges posed by modern technology.

Table of Contents

Literature review

As technology rapidly advances, cybercrimes such as internet phishing are increasingly prevalent, raising significant concerns about data privacy and security for users and enterprises worldwide. The widespread use of social networking sites, electronic healthcare systems, and online trading has generated vast amounts of data, contributing to big data. Analysing patient data brings up privacy and security concerns during data collection, storage, and processing. Consequently, there is a growing demand for Privacy Preserving Data Publishing (PPDP) to ensure secure data sharing over the internet (F. Li et al., 2011).

Privacy-preserving data sharing has garnered substantial attention owing to the sensitivity of identifiable data and the necessity to balance privacy concerns with the effective use of the shared information. To protect the privacy of data owners, various de-identification and anonymization techniques are applied before data is released to the public or used for secondary purposes (A. Mishra, 2024).

Differential privacy has become a key concept in this domain, offering a stringent framework for assessing privacy guarantees in data sharing process. The introduction of differential privacy by Dwork (Polat & Du, 2003) marked a significant milestone, providing a theoretical foundation for ensuring privacy guarantees while allowing data utility. Building on this foundation, subsequent research has explored the algorithmic basis of differential

privacy, detailing the mathematical complexities essential for creating effective privacy-preserving systems (Binjubeir et al., 2019).

Advances in technology have resulted in the creation of practical methods for privacy-preserving data sharing. Agrawal & Srikant, (2000) introduced randomized perturbation methods to maintain privacy in collaborative filtering systems, enabling personalized recommendations while protecting sensitive user data (Agrawal & Srikant, 2000). Jiang et al., (2019) focused on healthcare systems, investigating privacy-preserving methods tailored to address the unique challenges of sharing sensitive healthcare data among stakeholders (Jiang et al., 2019). Moreover, Sinjilawi et al., (2014) tackled privacy issues in cloud computing environments, providing methods to ensure secure data sharing within distributed computing infrastructures (Sinjilawi et al., 2014). Hai and Liu, (2022) designed privacy preserving data sharing for AI applications based on smart contracts, a blockchain-based method that transforms data sharing into model sharing, ensuring privacy and incentivizing data owners (Hai & Liu, 2022). In summary, these studies reveal the privacy-preserving data sharing requires a multifaceted strategy, incorporating foundational concepts such as differential privacy and utilizing practical techniques.

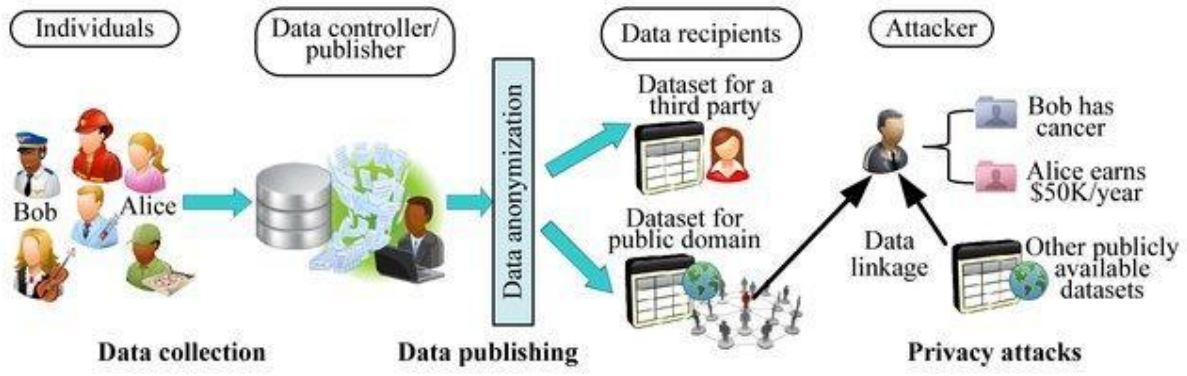
In this literature survey, the study is structured as follows: Section 1 delves into Privacy Preserving Data Publishing (PPDP). Section 2 explores various privacy-preserving techniques, focusing particularly on data anonymization methods like k-anonymity, generalization, suppression, masking, l-diversity, t-closeness, bucketization, randomization, and slicing detailing their principles, applications, limitations, and associated privacy preservation policies. Section 3 provides an in-depth discussion on Differential Privacy, including methods such as Data Perturbation. Section 4 outlines the privacy-preserving principles. Section 5 examines the ethical foundations of privacy, balancing principles in the technological era. Section 6 reviews related work and recent advances in privacy-preserving techniques. Section 7 discusses ISO/IEC standards for enhancing data protection. Section 8 presents the proposed framework for privacy preserving. Finally, Section 9 include the brief conclusion of the report.

1. Privacy Preserving Data Publishing

Privacy Preserving Data Publishing (PPDP) involves sanitizing personal data (modifying data to remove or obscure sensitive information), such as electronic health records, to make them

safe for release to agencies or the public. A typical scenario of PPDP is illustrated in Figure 1, outlining the various phases of data processing. An attacker could be any data recipient who gains access to an individual's personal information. Therefore, it is crucial for data publishers to implement various privacy-preserving techniques to control the information in the released data by modifying it prior to publication (Allard et al., 2014; Ayala-Rivera et al., 2014).

Figure 1. Outline of Privacy Preserving Data Publishing (PPDP)



Source: (Ayala-Rivera et al., 2014).

A variety of complex data mining algorithms have been developed to prevent data breaches and avoid penalties from government agencies. It is essential to implement techniques that ensure data privacy and facilitate safe information exchange at every stage of data mining (the process of discovering patterns and insights from large datasets, such as identifying fraudulent transactions in financial data or predicting patient outcomes in healthcare) (Qlik, 2024). Methods including access confinement (restricting access to only necessary data to enforce least privilege) and data distortion (introducing deviations from accurate data representation to prevent misuse) are used to protect sensitive information (Aldeen et al., 2015; Jain et al., 2016).

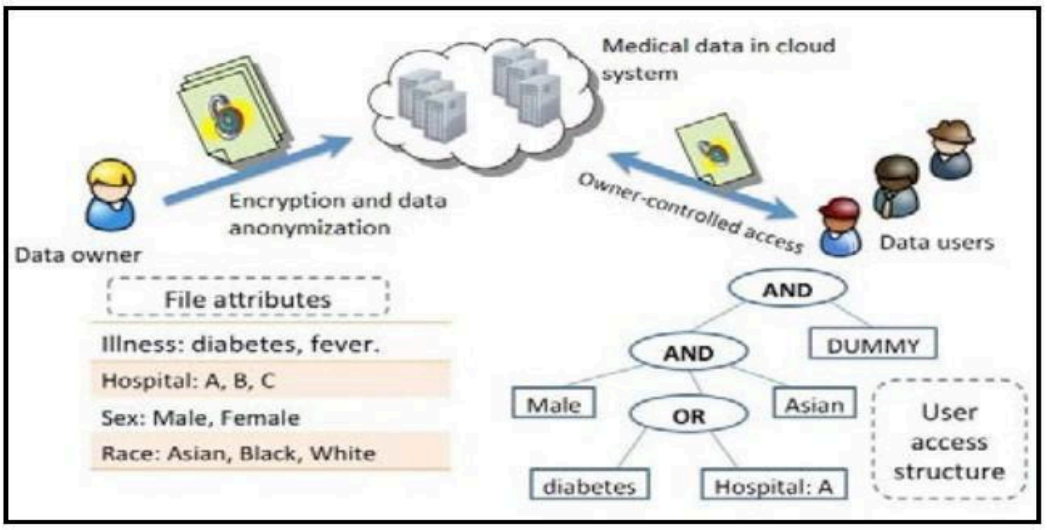
During the data storage phase, encryption techniques such as Identity-Based Encryption (IBE) (where the user's public key is derived from their identity) and Attribute-Based Encryption (ABE) (where access control is based on user attributes) are commonly used to secure data stored with cloud vendors or on medical servers (Ghafghazi et al., 2016). PPDP is primarily applied during the data processing stage of big data analytics. Data anonymization, also known as data masking or data desensitization, is employed to obscure or conceal sensitive information about individuals, thereby reducing the risk of re-identification (Emam, 2006; Vinogradov & Pastsyak, 2012).

PPDP employs anonymization techniques such as generalization and suppression to protect data by altering it to hide sensitive information. Furthermore, graph-based methods such as constrained perturbation are used to anonymize large data networks. These techniques include k-anonymity, l-diversity, classification, clustering, association rule, condensation, and cryptography (Aldeen et al., 2015; Casas-Roma et al., 2017; Jain et al., 2016).

For instance, medical data, which is high-dimensional and sourced from diverse origins, has been digitalized to reduce costs and improve quality, leading to the creation of Electronic Health Records (EHRs). EHRs store extensive patient information in centralized repositories, with cloud-based storage becoming increasingly popular. Due to the susceptibility of EHRs to cyber-attacks aimed at extracting personal information, anonymization techniques are essential to prevent data misuse and protect patient privacy (Nabeel et al., 2013). Figure 2 shows the anonymization of medical data.

Next section, in dept explain the various anonymization techniques have been developed to preserve the privacy of microdata, with three prominent methods being k-anonymity, l-diversity, and t-closeness. This review examines the advantages and disadvantages of these widely used techniques.

Figure 2. Anonymization of medical data



Source: (Nabeel et al., 2013).

2. Privacy Preserving Techniques

2.1 Data Anonymization

Data anonymization is essential for addressing privacy concerns. Anonymization focuses on converting personally identifiable information (PII) within datasets into a format that makes individual identification impossible. The process aims to protect individuals' privacy while allowing the data to be used for legitimate purposes such as research and analysis (Secoda, 2024). Anonymized data must exclude direct identifiers such as names, addresses, or social security numbers and be altered to prevent re-identification through linkage with other available data. This method enables hospitals, banks, social network providers, and insurance companies to share or publish user data while maintaining privacy, ensuring the data remains useful for legitimate purposes (Majeed & Lee, 2016).

Several techniques are used for data anonymization, including k-anonymity, generalization, masking, and t-closeness, each offering unique strengths and limitations. Each will be discussed in the following sections.

a) K-Anonymity

Many data holders, including governmental agencies and healthcare facilities, mistakenly believe that data, such as medical records, can retain anonymity if identifiable details such as names, addresses, and phone numbers are concealed before sharing the remaining records. However, the risk of re-identification remains when linking this data with other publicly available datasets, such as voter lists. While introducing noise into the dataset, such as through false values and scrambling, may provide anonymity, it can also lead to inaccurate statistical outcomes during data mining and analysis (Sweeney, 2002). To address these challenges, Samarati and Sweeney devised the k-anonymity technique in 1998, utilizing methods such as generalization and suppression to facilitate controlled data disclosure while safeguarding the integrity of tuple values. Quasi-identifiers, such as birth dates and genders, play a pivotal role in individual identification. A dataset achieves k-anonymity when each tuple's quasi-identifier values occur at least 'k' times, ensuring that tuples remain indistinguishable from one another (Samarati & Sweeney, 2007).

Principle of k-Anonymity:

K-anonymity as a privacy-preserving model designed to ensure that individual records within a dataset cannot be uniquely identified. A dataset is considered k-anonymous when each

record is indistinguishable from at least (k – 1) other records. This method effectively prevents the linkage of databases to re-identify individuals (Ravichandran et al., 2015). If every value within a dataset cannot be distinguished from at least (k-1) other records within the same table, it meets the criterion of k-anonymity. A higher value of k indicates stronger privacy protection (Hussien et al., 2013).

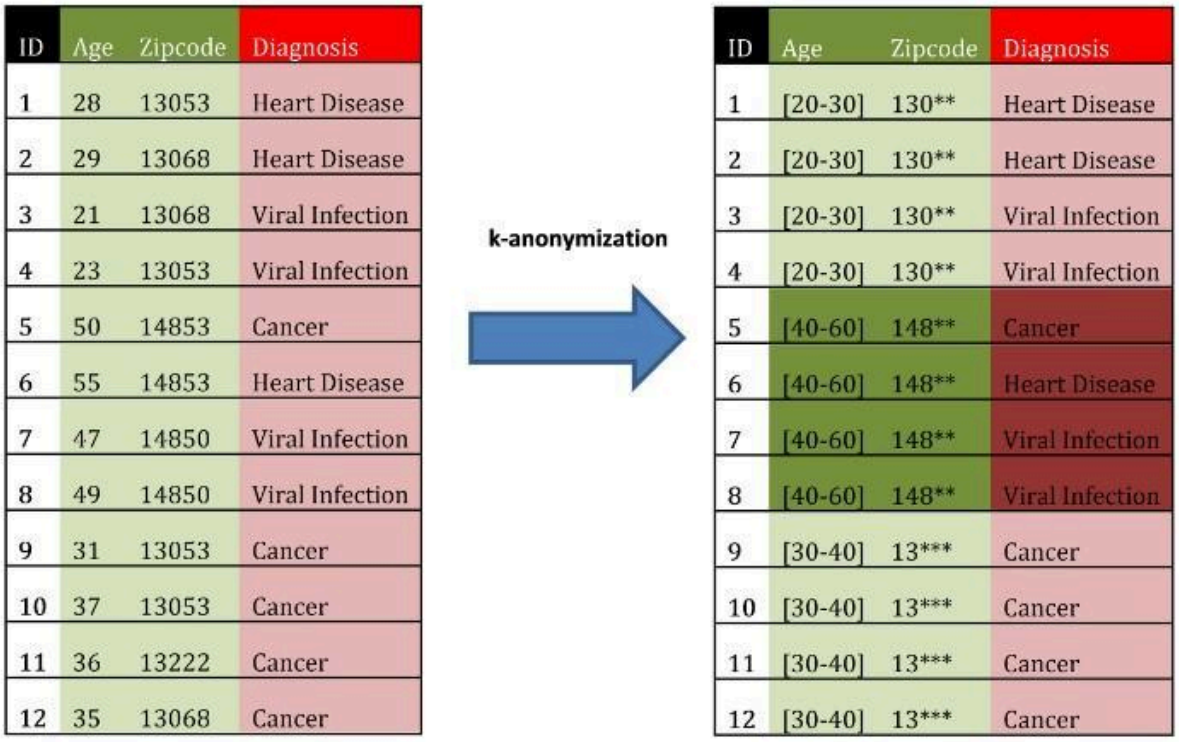
Table 1 illustrates an example of the k-Anonymity process. A release provides k-anonymity protection if the information for each individual in the release cannot be distinguished from at least k-1 other individuals whose information is also included. This means that a dataset is k-anonymous as long as there are at least k entries for any given set of quasi-identifiers. For instance, if the quasi-identifiers are zip code, gender, and age, the dataset would be 3-anonymous if every possible combination of these attributes has at least 3 entries.

This helps prevent linkage attacks, as an attacker cannot link the data to another database with high certainty. However, it does not protect against homogeneity attacks, where all k individuals have the same value for an attribute. For instance, if three individuals in the same age group, gender, and zip code also all have the same type of cancer, their privacy is not protected by k-anonymity. Similarly, k-anonymity does not defend against attacks that leverage background knowledge. Table 2 shows an example of homogeneity and background knowledge attacks on k-anonymous datasets.

Pros of k-Anonymity:

- k-Anonymity protects against identity disclosure by preventing links to datasets with fewer than 'k' values. The blocks adversaries from connecting sensitive data with external datasets (Hussien et al., 2013).
- The cost of implementing the method is significantly lower compared to other anonymity methods, such as cryptographic solutions (Belsis & Pantziou, 2012).
- Algorithms including Datafly, Incognito, and Mondrian are widely used in k-anonymity, especially in Privacy-Preserving Data Publishing (PPDP). In addition, clustering is integrated into k-anonymity to improve privacy preservation (Ayala-Rivera et al., 2014; Rana & Jayabalan, 2016).

Table 1. k-anonymity Example



Source: (Unal et al., 2022).

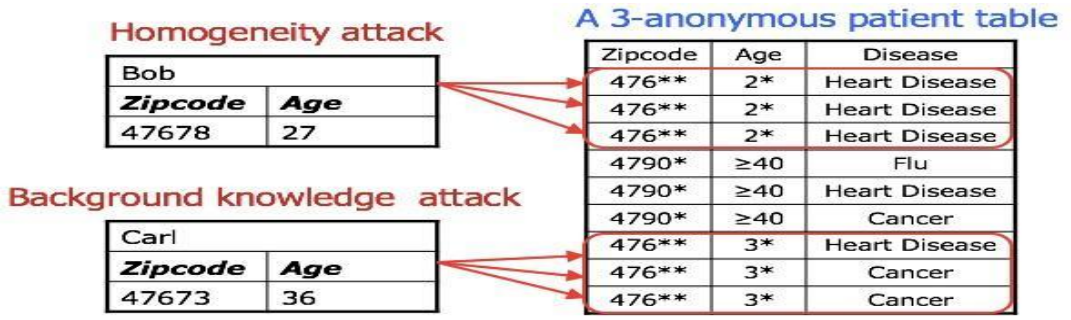
Cons of k-Anonymity:

The k-Anonymity technique has several recognized limitations, primarily involving attacks such as unsorted matching, complementary release, minimality, and temporal attacks (Hussien et al., 2013; Jain et al., 2016; Sweeney, 2002). Furthermore, the k-Anonymity technique can lead to significant utility loss when applied to high-dimensional data, and special measures are required if the data has already been anonymized more than once (Singh, 2023). The following sections briefly explain two well-known attacks on k-anonymity.

- Homogeneity attack: A homogeneity attack occurs when there is insufficient diversity in sensitive attributes, leading to clusters that can reveal information. For example, if A and B are adversaries, and A knows B's zip code and age, A can potentially determine B's medical condition. If all medical records matching B's information show the same condition (e.g., cancer), A can infer B's health status. To prevent this, the k-anonymous table should be further refined by diversifying the sensitive values among tuples with similar quasi-identifiers (Hussien et al., 2013; Machanavajjhala et al., 2007).

- Background knowledge attack: In a background knowledge attack, the adversary uses prior knowledge about an individual and logical reasoning to uncover sensitive attributes. For example, if A and C are acquaintances, and A wants to infer C's personal data from the same patient record as B, A can use known information about C. Suppose A knows that C is a 45-year-old Asian female living in a specific zip code. The record indicates that C could have one of three diseases: cancer, heart disease, or viral infection. With additional knowledge that C avoids high-calorie meals and has low blood pressure, A concludes that C likely has heart disease. Thus, k-anonymity is vulnerable to background knowledge attacks (Hussien et al., 2013; Machanavajjhala et al., 2007).

Table 2. Homogeneity and Background Knowledge Attacks on k-Anonymous Data



Source: (oananiculaescu, 2017).

b) Generalization

Generalization is a common anonymization technique entails substituting specific values in the data with more generalized ones. For instance, exact ages can be replaced with age ranges (e.g., 20-30 years old), or precise geographic locations can be replaced with broader areas (e.g., substituting street addresses with city names) (A. Mishra, 2024). In other word, generalization approach replaces quasi-identifier values with less specific but semantically consistent values.

In this method, all quasi-identifier values within a group are generalized to cover the entire group's range in the QID space (Ghinita et al., 2011). If at least two transactions in a group have different values in a particular column (i.e., one includes an item and the other does not), all information about that item in the current group is lost. The QID used in this process encompasses all possible items in the log. Given the high dimensionality of the

quasi-identifier, with thousands of possible items, any generalization method would likely result in very high information loss, making the data unusable (Ghinita et al., 2008).

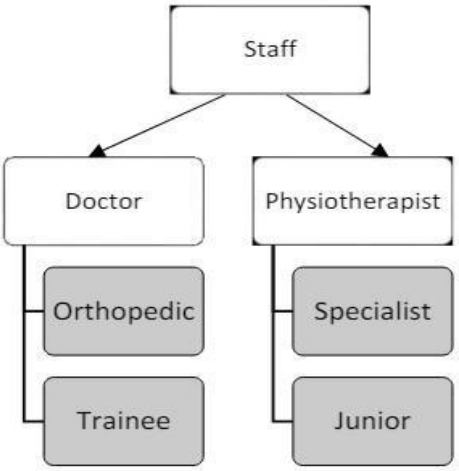
For generalization to be effective, records within the same bucket must be sufficiently similar, ensuring that generalizing these records does not result in substantial information loss. However, in high-dimensional data, the majority of data points have comparable distances from each other. When performing data analysis or data mining tasks on the generalized table, the data analyst must assume a uniform distribution, where every value within a generalized interval or set is equally probable, as no alternative distribution assumption can be justified. This significantly reduces the utility of the generalized data. Furthermore, because each attribute is generalized independently, correlations between different attributes are lost. To examine attribute correlations on the generalized table, the data analyst must assume that every possible combination of attribute values is equally probable. This inherent issue with generalization impedes the effective analysis of attribute correlations (Mandal & Nigam, 2012).

The generalization technique can be categorized into four different schemes (Rana & Jayabalan, 2016). Each scheme will be explained in detail using the example provided in Figure 3.

1) Full Domain Generalization

The Full Domain Generalization approach results in the greatest distortion of the dataset. The quasi-identifier value is generalized to the same level within a predefined hierarchical structure. Consequently, if the node "orthopedic" is generalized to its parent node "Doctor," all corresponding child nodes under "Doctor" must also be generalized.

Figure 3. Hospital Staff Tree



Source: (Rana & Jayabalan, 2016).

2) Subtree Generalization

Compared to full domain generalization, sub-tree generalization results in less distortion to the original dataset. When a specific node in the taxonomy tree is generalized to its parent node, all the child nodes of that specific tree must be generalized to the parent node, but other nodes remain unaffected. For example, generalizing the "Doctor" node and its sub-nodes does not require generalizing the "Physiotherapist" node and its sub-nodes. Another variant, known as "Unrestricted Subtree Generalization," follows a similar concept but allows for more granular data masking. For instance, the "Orthopedic" node can be generalized under the "Doctor" node without needing to generalize the "Trainee" node.

3) Cell generalization

Cell generalization results in the least data distortion compared to other generalization techniques. This method focuses on individual records rather than the entire dataset. For instance, when the "Orthopedic" node is generalized to its parent node "Doctor," other records within the "Doctor" node do not need to be generalized.

4) Multidimensional Generalization

This method generalizes data by grouping various values that fall under quasi-identifiers. For example, "[Orthopedic, Female]" can be generalized to "[Doctor, Gender]," while "[Trainee, Male]" can be generalized to "[Doctor, Male]." In essence, this technique generalizes records based on combinations of quasi-identifiers (Levich et al., 2010).

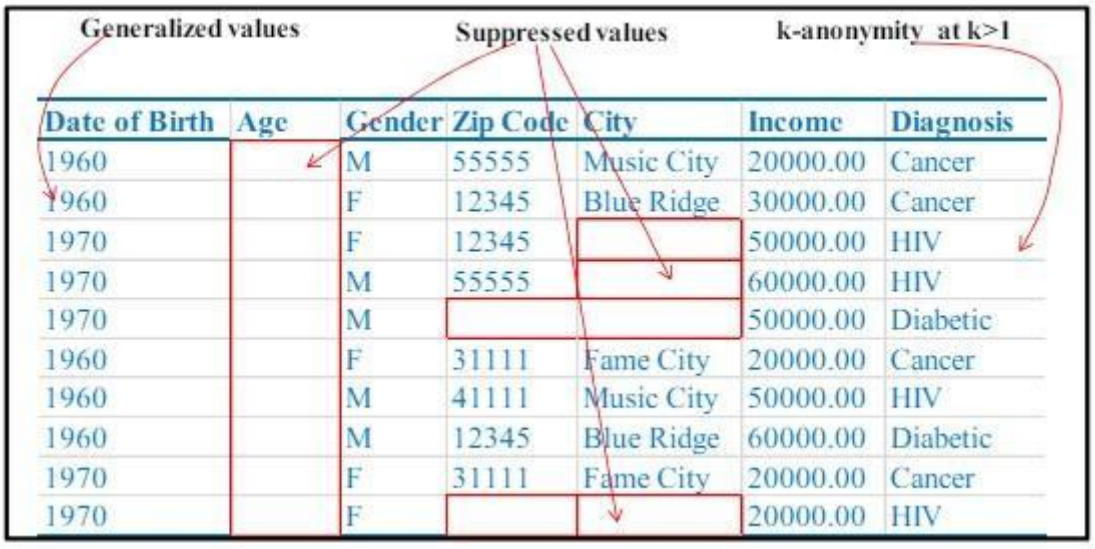
c) Suppression

Suppression, in conjunction with generalization to bolster k-anonymity. Suppression, also known as data masking or redaction, involves concealing specific values within the quasi-identifiers (Samarati & Sweeney, 2007). These masked values are denoted by an asterisk (*) and can be applied across both domain and value generalization hierarchies. Referring back to the example discussed in the generalization section, the mapped value can be initially suppressed as Z1 (0212*, 0212*) and subsequently further suppressed to Z2 (021**), eventually achieving maximum suppression (Sweeney, 2002). Figure 4 visualizes the concepts of generalization, suppression, and k-anonymity.

Suppression can be applied at different levels to protect sensitive information in a dataset. At the tuple level (TS), suppression means that an entire tuple, or row of data, is suppressed. At the attribute level (AS), suppression involves removing an entire attribute, or column of data, from the dataset. At the cell level (CS), individual cells within the dataset can be suppressed independently of others, allowing for more granular control over which specific pieces of data are removed. Each method of suppression helps to enhance privacy by reducing the risk of re-identifying individuals within the dataset (Samarati, 2011).

This technique is especially valuable for handling highly sensitive information that cannot be anonymized by other methods. For instance, in a financial dataset, names and account numbers may be suppressed while retaining other, less sensitive attributes (Faster Capital, 2024).

Figure 4. Visualization of Generalization, Suppression, and k-Anonymity



Source: (Mivule, 2014).

d) Masking

Masking involves substituting sensitive information with other values that maintain the original data's format but do not disclose sensitive details. For instance, this could mean replacing the last few digits of a credit card number with asterisks or substituting names with generic labels or you can replace a value character with a symbol such as "*" or "x" (Imperva, 2024; A. Mishra, 2024). Masking data stands out as a highly secure method for anonymizing data, ensuring that the original data remains unidentifiable and cannot be reversed. This approach is frequently employed to adhere to consumer privacy regulations and to conceal sensitive information such as financial data, Protected Health Information (PHI), and intellectual property (Richman, 2023). Data masking can take various forms depending on the techniques employed. Figure 5 shows some examples of the different types.

Figure 5. Examples of data masking techniques

Source: (Cobb, 2024).

Masking can be categorized into non-perturbative and perturbative methods. Non-perturbative masking reduces the level of detail in the original data without distorting it. Common non-perturbative techniques include publishing only a sample of the original dataset (Sampling), coarsening categorical attributes by combining categories into more general ones (Generalization), replacing numerical values with intervals, grouping values above or below a certain threshold into a single category (Top/Bottom coding), or suppressing certain attribute values to reduce individual uniqueness (Local suppression) (Salas & Domingo-Ferrer, 2018).

Some perturbative masking techniques such as noise addition, rank swapping, post-randomization and micro aggregation are special cases of matrix masking (Duncan & Pearson, 1991), where the original microdata set is X and the masked microdata set is Z

computed as Z = AXB + C, matrix A is called a record-transforming mask, matrix B an attribute transforming mask and matrix C a displacing mask (Klein et al., 2002).

e) l-Diversity

l-Diversity was introduced to address the limitations of k-anonymity. It extends k-anonymity by ensuring data privacy without needing to know the adversary's background knowledge, thus preventing attribute disclosure. This method focuses on ensuring that the sensitive attributes within each group are "well-represented." Essentially, it modifies k-anonymity by incorporating its principles while enhancing the diversity of sensitive attributes within each group (N. Li et al., 2007; Machanavajjhala et al., 2007). For instance, if there are 4 types of disease diagnoses, a 2-diverse equivalence class would ensure that the sensitive attribute appears in no more than two out of the four possible records. Figure 6 represents the visualization of a 2-diverse table.

Figure 6. Visualization of a 2-Diverse Table

PatientName	Postcode	AgeGroup	Disease	
*	476**	Under 30	Heart Disease	$\{QI\ Group\ 1\}$
*	476**	Under 30	Heart Disease	
*	476**	Under 30	Flu	
*	476**	Under 30	Cancer	
*	479**	Over 40	Heart Disease	$\{QI\ Group\ 2\}$
*	479**	Over 40	Heart Disease	
*	479**	Over 40	Flu	
*	479**	Over 40	Pneumonia	

Source: (Information with Insight, 2024).

Principle of l-Diversity:

A k-anonymous table is considered l-diverse if each equivalence class contains at least 'l' "well-represented" values for each sensitive attribute (Casas-Roma et al., 2017; Machanavajjhala et al., 2007). "Well-represented" can be understood through the following principles:

- Distinct l-diversity: Each value appears more frequently than other values within the equivalence class. However, this could allow an attacker to infer the likely value based on its higher frequency.

- Entropy l-diversity: The table must have an entropy of at least log(l) to satisfy entropy l-diversity for each equivalence class. This method may be restrictive if the overall table entropy is low, with only a few recurring values.
- Recursive (c, l)-diversity: This principle requires that sensitive values in each equivalence class neither occur too frequently nor too rarely. It is a stronger notion compared to the previous two principles (N. Li et al., 2007; Machanavajjhala et al., 2007).

Pros of l-Diversity:

- Ensures a broader distribution of sensitive attributes within each group, enhancing data protection.
- Offers improved defence against attribute disclosure, building upon the k-anonymity technique.
- l-Diversity performs slightly better than k-anonymity due to the more efficient pruning capabilities of the l-diversity algorithm (Casas-Roma et al., 2017; N. Li et al., 2007; Machanavajjhala et al., 2007).

Cons of l-Diversity:

- Achieving l-diversity can be redundant and labour-intensive.
- It is vulnerable to skewness and similarity attacks, as it may not sufficiently prevent attribute exposure due to the semantic relationships between sensitive attributes (Jain et al., 2016; N. Li et al., 2007).

f) T-Closeness

An improvement over l-diversity is the t-closeness technique, which reduces the granularity of the interpreted data. This technique limits the observer's knowledge of specific data while not restricting their understanding of the overall dataset. As a result, it diminishes the correlation between quasi-identifier attributes and sensitive attributes. The distance between distributions is measured using Earth Mover's Distance (EMD). For categorical attributes, EMD measures the distance between values based on the minimum level of generalization within the domain hierarchy (N. Li et al., 2007).

T-closeness aims to maintain a similarity between the distribution of sensitive attribute values within each equivalence class (a group of records with identical quasi-identifiers) and the overall distribution in the dataset. This prevents attackers from

deducing sensitive information by analyzing statistical outliers within equivalence classes (A. Mishra, 2024).

T-closeness is introduced as a privacy-preserving approach to overcome the limitations of existing methods like k-anonymity and l-diversity. In l-diversity, a concern arises where adversaries can infer sensitive attribute information if they know the distribution of those attributes, posing a significant drawback. Moreover, many privacy methods assume attributes to have categorical values, leaving them vulnerable to attacks such as distribution skewness and semantic similarity within equivalence classes.

T-closeness addresses the related issue by ensuring that the distance between the sensitive attribute of an equivalence class and that of the entire dataset is within a defined threshold, known as t. This helps mitigate the risk of adversaries inferring unique individual information. The Earth Mover's Distance (EMD) metric measures the distribution distance between sensitive attributes, considering the semantic proximity of attribute values. While t-closeness preserves feature disclosure, identity disclosure remains a concern. Hence, a combination of k-anonymity and t-closeness can effectively safeguard the privacy of published data (N. Li et al., 2007).

Pros of t-Closeness:

- It prevents attribute disclosure, thereby safeguarding data privacy.
- Protects against homogeneity and background knowledge attacks that are vulnerabilities of k-anonymity.
- Recognizes the semantic closeness of attributes, addressing a limitation of l-diversity.

Cons of t-Closeness:

- Utilizing Earth Mover’s Distance (EMD) in t-closeness makes it challenging to determine the closeness between the t-value and the knowledge acquired.
- Requires that the distribution of sensitive attributes within each equivalence class be similar to that in the overall table (Jain et al., 2016; N. Li et al., 2007).

g) Bucketization

Bucketization is an anonymization technique used for publishing sensitive data. It works by grouping records into small buckets to hide the direct association between sensitive information and identifying information at the record level. Unlike the traditional

generalization technique, bucketization does not require a taxonomy of attribute values, making it suitable for a wider range of dataset (K. Wang et al., 2016).

Generalization converts QI attribute values into more general forms, grouping tuples with the same generalized values into equivalence groups. Consequently, records within the same equivalence group become indistinguishable. In contrast, bucketization splits tuples into buckets, disrupting the link between QI attributes and sensitive attributes. As a result, each record corresponds to various sensitive values within its bucket (B. Li & He, 2023).

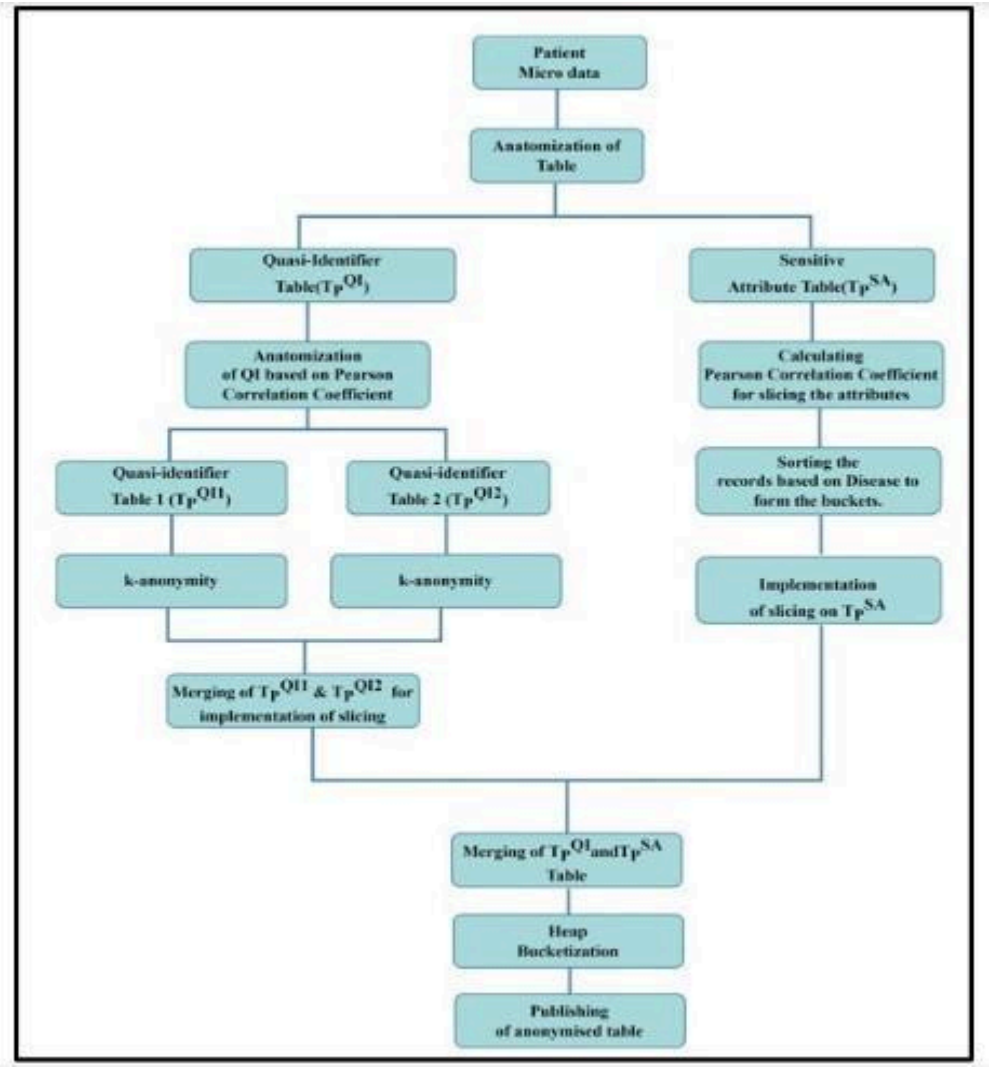
The efficient Heap Bucketization-anonymity (HBA) model has been introduced to balance privacy and utility when dealing with multiple sensitive attributes. The main objective of the HB-anonymity algorithm is to balance privacy and utility. The Heap Bucketization technique is specifically designed to address the shortcomings of traditional bucketization. The HB-anonymity model integrates generalization, slicing, and bucketization methods. This model employs anatomization to vertically partition the dataset into two tables: one for quasi-identifiers and another for sensitive attributes (Jayapradha et al., 2022).

In the HB-anonymity algorithm, as Figure 7 shows, the patient table and a variable k are inputted in line 1, resulting in a heap bucketized table. The patient table is divided into a quasi-identifier table and a sensitive attribute table in lines 3 and 4. Pearson correlation is used to calculate correlations among quasi-identifiers in lines 5 and 6. The k variable is used to anonymize the quasi-identifier and correlation table in line 7, and further anatomization occurs in line 8. Lines 9 and 10 apply k-anonymity to the two quasi-identifier tables.

After merging these tables, slicing is applied to highly correlated attributes in lines 11 and 12. Correlations among sensitive attributes are calculated in lines 13 and 14, and the sensitive attribute table is anonymized based on these correlations in line 15. Blood pressure is split into sys and dys fields in line 16, and the sensitive attribute table includes Temp, Pulse Rate, Respiratory Rate, Sys, Dys, and Disease in line 17.

Attributes are sorted by disease in line 18, and buckets with three records each are formed from lines 19 to 23. Slicing is performed on the bucketized table in line 24, and the tables are merged and sorted by bucket id in lines 25 and 26. Finally, from lines 27 to 29, values of highly correlated attributes are grouped in each bucket, and records are sorted by id for data publishing (Jayapradha et al., 2022).

Figure 7. Workflow of the HB-anonymity



Source: (Jayapradha et al., 2022).

h) Randomization

Randomization is a commonly used technique that adds noise or perturbations to a dataset to prevent the re-identification of individuals. This can be done by inserting random values into the original data, such as adding random noise to numerical attributes or substituting specific identifiers with random values. For instance, in a healthcare dataset, the ages of patients could be altered by adding a random number within a specified range to protect their privacy.

As explained by Pingshui Wang in (Kumar Singh et al., 2014), the randomization method generally seeks to find an appropriate balance between preserving privacy and enabling knowledge discovery. Representative randomization techniques include random-noise-based perturbation and the Randomized Response scheme.

Randomization is used in surveys, sentiment analysis, and similar applications. This technique does not require knowledge of other records in the dataset and can be implemented during data collection and preprocessing. Randomization involves no anonymization overhead. However, its application to large datasets is impractical due to time complexity and data utility issues (Ram Mohan Rao et al., 2018).

i) Slicing

The new privacy preservation technique, slicing, offers more security compared to other methods. Slicing partitions the data both horizontally and vertically. Vertical partitioning groups attributes into columns based on their correlations (Kumar Singh et al., 2014). This method is more secure than generalization and bucketization due to its comprehensive process, which involves the following functional steps:

1) Extract the dataset from the database.
2) Divide the records during the anonymity process.
3) Interchange the sensitive values.
4) Generate and display multi-set values.
5) Correlate attributes and display secure data (T. Li et al., 2012).

Table 3 represents the original data, including the three quasi-identifier (QI) attributes: Age, Sex, and Zipcode, with the sensitive attribute (SA) being Disease. Subsequently, Table 4 displays the sliced data. In the first bucket of the sliced table, the values {(22, M), (22, F), (33, F), (52, F)} are randomly permuted, as are the values {(47906, dyspepsia), (47906, flu), (47905, flu), (47905, bronchitis)}, to obscure the direct linkage between the two columns within the bucket (Bhuvanesh et al., 2017).

Comparison with Generalization:

Generalization includes several types of recoding, with local recoding preserving the most information. In local recoding, tuples are grouped into buckets, and for each bucket, all values of one attribute are replaced with a generalized value. This approach is local because the same attribute value may be generalized differently in different buckets (Bhuvanesh et al., 2017).

Comparison with Bucketization:

Slicing can be compared to bucketization by noting that bucketization is a specific form of slicing, where there are only two columns: one for the sensitive attribute (SA) and one for all

quasi-identifiers (QIs) (Bhuvanesh et al., 2017). Slicing has several advantages over bucketization:

- Prevention of Membership Disclosure: Slicing can partition attributes into more than two columns, helping to prevent membership disclosure, whereas bucketization does not.
- Flexible Attribute Separation: Slicing does not require a strict separation of QIs and the SA. This is useful for datasets like census data, where it's difficult to separate QIs from SAs due to the lack of a single external database for reference.
- Preservation of Attribute Correlations: Slicing allows columns to contain both QI attributes and the SA, preserving correlations between them. For example, having Zipcode and Disease in one column allows for the analysis of their relationship, which is important for data utility. For workloads that consider attributes in isolation, one can simply publish two tables: one with all QIs and one with the SA.

Table 3. The Original

Age	Sex	Zip code	Disease
22	M	47906	dyspepsia
22	F	47906	flu
33	F	47905	flu
52	F	47905	bronchitis
54	M	47302	flu
60	M	47302	dyspepsia
60	M	47304	dyspepsia
64	F	47304	gastritis

Source: (Bhuvanesh et al., 2017).

Table 4. The Sliced

(Age, Sex)	(Zip code, Disease)
(22, M) (22, F) (33, F) (52, F)	(47905,flu) (47906,dysp.) (47905,bron.) (47906,flu)
(54, M) (60, M) (60, M) (64, F)	(47304,gast.) (47302,flu) (47302,dysp.) (47304,dysp.)

Source: (Bhuvanesh et al., 2017).

3. Differential Privacy

Overview:

Differential privacy, introduced by Dwork et al. in 2006, has become a key concept in privacy-preserving data analysis. It provides a robust framework for measuring the privacy guarantees of algorithms handling sensitive information (Dwork et al., 2006). Essentially, differential privacy ensures that the outcome of a computation is minimally impacted by the inclusion or exclusion of any individual's data. This means that it is difficult for an observer to determine whether a specific individual's data is part of the dataset based on the computation's results (Dwork & Roth, 2013).

The core principle of differential privacy is "indistinguishability": an algorithm that meets this standard produces a probability distribution of outputs that is nearly the same whether or not a particular individual's data is included. This feature enables the sharing of aggregated or analysed data without compromising the privacy of individual contributors (Dwork et al., 2006).

Privacy Guarantees:

Differential privacy provides robust mathematical assurances for individual privacy protection. It guarantees that an individual's data has a minimal effect on the overall result of a computation, thereby preventing adversaries from deducing sensitive information about individuals from the results. The privacy level of a differentially private algorithm is measured by a parameter called ε (epsilon). This parameter indicates the maximum permitted difference in the likelihood of different outcomes due to the presence or absence of an individual's data. A lower ε value signifies a stronger privacy guarantee, as it means the output distribution changes less with any individual's data contribution (Hay et al., 2016).

Data Perturbation:

Data perturbation methods are integral to implementing differential privacy. Data perturbation is a method employed in privacy-preserving data mining to safeguard sensitive information. Data perturbation is vital in privacy-preserving data sharing platforms, as it involves adding controlled noise or modifications to the original dataset. The main goal of data perturbation is to safeguard individuals' sensitive information while preserving the data's usefulness for analysis and decision-making (Wilson & Rosen, 2023).

Perturbation Methods:

Several techniques are employed for data perturbation, each tailored to different data types and privacy needs: Additive Perturbation, Multiplicative Perturbation, K-anonymity, Data Swapping, Micro-aggregation, Resampling, and Data Shuffling.

- Additive Perturbation involves masking attribute values by adding noise to the original data, making it impossible to recover individual records (Sriharsha & Parthasarathy, 2015).
- Multiplicative perturbations can also be effective for privacy-preserving data mining, conserving inter-record distances and thus allowing altered records to be used in conjunction with distance-intensive data mining applications. Unlike additive perturbation, multiplicative perturbation preserves more sensitive information such as distances (Rahman et al., 2023).
- K-anonymity comprises two methods: Generalization and suppression. Generalization involves generalizing attribute values, for example, transforming a date of birth into just the year of birth. Suppression involves completely removing attribute values, reducing the risk of recognition while potentially decreasing the accuracy of applications using the modified data (Slijepčević et al., 2021).
- Data Swapping preserves confidentiality in datasets with categorical variables by replacing values of sensitive variables between individual records (Pika et al., 2020).
- Micro-aggregation clusters data into small groups before publication, with each group's common value replacing each individual value (Kokolakis & Fouskakis, 2009).

Secure Multi-Party Computation (SMPC)

Secure Multi-Party Computation (SMPC) is a cryptographic method that facilitates multiple parties to collectively calculate a function using their individual inputs, all without disclosing those inputs to one another. SMPC aligns with the principles of differential privacy by allowing multiple parties to perform computations collaboratively without revealing their individual inputs, thus preserving data privacy.

The core aim of SMPC is to maintain privacy and confidentiality while enabling collaboration among parties in computations. This is accomplished by breaking down the computation into smaller segments, which each party executes independently, and subsequently merging the outcomes in a manner that conceals the inputs (Zhao et al., 2019).

Applications:

a. Secure Data Aggregation

SMPC is widely used in secure data aggregation contexts, particularly when multiple parties seek to merge their data for analysis without revealing individual data points. For instance, in healthcare settings, various hospitals may seek to aggregate patient data for statistical analysis in research endeavours while safeguarding patient privacy. SMPC enables these hospitals to calculate aggregate statistics, like averages or sums, without divulging sensitive patient details (Torkzadehmahani et al., 2022).

b. Collaborative Analytics

SMPC is widely used in secure data aggregation contexts, particularly when multiple parties seek to merge their data for analysis without revealing individual data points. For instance, in healthcare settings, various hospitals may seek to aggregate patient data for statistical analysis in research endeavours while safeguarding patient privacy. SMPC enables these hospitals to calculate aggregate statistics, like averages or sums, without divulging sensitive patient details (Torkzadehmahani et al., 2022).

c. Secure Outsourcing of Computations

SMPC also enables the secure delegation of computations to untrusted third parties. Entities can assign computationally intensive tasks to cloud service providers while guaranteeing the confidentiality of their data. Through SMPC protocols, parties can validate the accuracy of computation results without disclosing their inputs or intermediate values to the cloud provider (Information Commissioner’s Office, 2022).

4. Privacy-Preserving Principles:

Understanding privacy-preserving principles is crucial in today's digital landscape for several compelling reasons. First and foremost, these principles are vital for safeguarding personal data against unauthorized access, use, or disclosure. By adhering to privacy-preserving guidelines, organizations can ensure that individuals' rights to privacy are respected throughout the data lifecycle. Moreover, compliance with these principles is essential for meeting legal and regulatory requirements governing data protection. Laws such as the GDPR in Europe and similar regulations worldwide require organizations to implement robust measures to protect personal data, failure of which can result in substantial fines and

reputational damage. Building trust with users, customers, and stakeholders is another critical outcome of following privacy-preserving principles. When individuals feel confident that their data is handled securely and ethically, they are more likely to engage with organizations and share their information. Furthermore, understanding these principles helps mitigate risks associated with data breaches and other security incidents, thereby safeguarding both individuals and organizations. Ultimately, embracing privacy-preserving principles not only ensures legal compliance but also fosters a culture of ethical data stewardship, reinforcing trust and accountability in the digital age.

This section covered fundamental principles found in global data protection laws and frameworks like the GDPR in Europe, as well as similar regulations in other jurisdictions. These principles provide guidance to organizations on responsibly managing personal data while upholding individuals' privacy rights.

1. **Consent and Control**

In the realm of safeguarding privacy in data practices, "Consent and Control" embodies principles and mechanisms designed to empower individuals with authority over the collection, utilization, and dissemination of their personal data. This principle underscores the importance of informing individuals about organizational data practices, providing them with the choice to consent to or decline the collection and processing of their data, and enabling them to manage their preferences over time.

Consent-Based Access Control (CBAC):

Zhang et al. (2016) propose a consent-based access control (CBAC) mechanism for healthcare systems to bolster security and privacy preserving in information exchange. This method involves patients granting data access to specific requesters through mutual agreements. Once consent is given, healthcare organizations can access encrypted data provided by healthcare providers, secured using conditional proxy re-encryption, which allows data encrypted for one recipient to be transformed securely so that it can be decrypted by another recipient under specific conditions. The scheme ensures protection against unauthorized access, including by public data centers, and achieves collusion resistance, mutual authentication, and contextual privacy. Evaluations demonstrate its high computational efficiency (A. Zhang et al., 2016).

Public Key-Based Consent Management:

The Public Key-Based Consent Management system provides a robust framework for ensuring privacy and data integrity in research contexts. Leveraging digital signatures, participants can securely declare their consent without revealing identifiable information. Each consent declaration, signed with the participant's private key, integrates seamlessly into the record database, ensuring transparent and accountable data handling practices. This approach safeguards participant privacy by anonymizing consent declarations and establishing an auditable trail to verify lawful data usage (Jonas et al., 2019).

Moreover, the system supports participants' rights to withdraw consent and request data deletion. When a participant initiates a delete statement, it is treated as a standard record and linked with all pertinent data, including the initial consent declaration. This linkage ensures that all participant data associated with the withdrawal request are marked for deletion. Subsequently, a garbage collection mechanism systematically cleanses the database of flagged records. Concurrently, public keys linked to deleted data groups are removed from the public key store to prevent unauthorized access. This comprehensive methodology not only meets regulatory standards but also enhances transparency and participant trust by guaranteeing that withdrawn consent leads to the complete removal of associated data (Jonas et al., 2019).

All operations related to record grouping and consent management can be implemented as a MapReduce problem using big data technologies like Apache Spark or Hadoop. This approach ensures scalability and efficiency in handling large datasets, thereby supporting robust data governance and compliance with privacy regulations (Jonas et al., 2019).

2. **Purpose Limitation**

Personal data must be processed for "specified, explicit, and legitimate purposes," meaning that the purpose of processing must be clearly defined before processing begins and upheld throughout the entire lifecycle of personal data. According to the GDPR (recital 33) (GDPR, 2024b), there are instances, particularly in scientific research, where it may not be feasible to fully identify the purposes of data processing at the time of collection. Therefore, individuals should have the opportunity to consent to specific areas of scientific research in accordance with recognized ethical standards. However, this does not imply that a broad purpose definition (e.g., "for language research") suffices; research purposes should be well-described at the time of collection and may evolve over time (Clarin, 2024; Von Grafenstein, 2018).

Once the purpose of data processing is established, processing for a purpose incompatible with the original one is prohibited. Conversely, data may be processed for a new, different purpose if this new purpose aligns with the original purpose. As an exception, the GDPR specifies that scientific research is always considered a compatible purpose. This means that data lawfully collected for any purpose (e.g., statistics or accounting) can be reused for scientific research purposes, provided that all other data protection principles are adhered to (Clarin, 2024; ITA, 2024).

The purpose limitation principle, a fundamental aspect of data privacy regulations, governs how personal data can be utilized. It mandates that personal data must be collected and processed for clearly defined, explicit, and lawful purposes, and cannot be further processed in a manner that is incompatible with those original purposes. This principle ensures that personal data collected and managed within cloud environments is strictly used for specific and legitimate reasons (Palo Alto Networks, 2024).

Organizations are required to articulate the purpose for collecting personal data and refrain from any subsequent processing that does not align with the original intent. By adhering to this principle, organizations can achieve compliance with data privacy laws, minimize the risk of unauthorized data usage, and uphold the trust of their customers and stakeholders (Healey, 2024).

To maintain the intended purposes of data collection, Sultan and Jensen, (2021) proposal suggests incorporating data into immutable and append-only resource metadata (provenance), accessible through an access control mechanism for verifying data usage. This approach not only upholds purpose limitation in extensive infrastructures but also enhances transparency for individuals and auditors to monitor the utilization of personal information (Sultan & Jensen, 2021).

Principle of Data Integrity and Purpose Limitation:

a. Consistent with the Principles, personal information must be limited to the information that is relevant for the purposes of processing. An organization may not process personal information in a way that is incompatible with the purposes for which it has been collected or subsequently authorized by the individual. To the extent necessary for those purposes, an organization must take reasonable steps to ensure that personal data is reliable for its intended use, accurate, complete, and current. An organization must adhere to the Principles for as long as it retains such information.

b. Personal data can only be kept in an identifiable form for as long as it is needed for processing. However, this does not prevent organizations from retaining data for longer periods if it serves purposes like archiving in the public interest, journalism, literature and art, scientific or historical research, and statistical analysis. In such cases, the processing must still comply with the other principles and provisions of the EU-U.S. DPF, and organizations should take appropriate measures to ensure compliance.
c. Depending on the situation, compatible processing purposes may include activities that support customer relations, compliance with legal requirements, auditing, security and fraud prevention, and protecting or defending the organization's legal rights. Other purposes that align with the expectations of a reasonable person in the context of data collection are also considered compatible.
d. In this context, if the identification methods reasonably likely to be used (taking into account factors like cost, time required for identification, and available technology) and the format in which the data is kept make it possible for the organization or a third party with access to identify an individual, then the individual is considered identifiable (ITA, 2024; Von Grafenstein, 2018).

3. **Data Minimization**

Data minimization is a direct way to limit privacy leakage (Qu et al., 2019). Data minimization dictates that personal data should be "adequate, relevant, and limited to what is necessary for the purposes of processing." This principle requires that only data essential to achieving the intended purpose can be lawfully collected, stored, or processed. It poses a challenge in data-intensive operations where every piece of data may be valuable but not all are essential for the intended purpose (Clarin, 2024).

Data minimization encourages organizations to carefully select the personal information they collect and process, aiming to mitigate privacy and regulatory risks associated with managing excessive data volumes (Qu et al., 2019). Understanding the importance of data minimization is the first step toward a more secure, ethical, and privacy-conscious digital landscape (Dabah, 2023). Here are the primary reasons why data minimization is crucial for responsible data management:

1. Reducing Privacy Risk: Data minimization significantly lowers the risk of data exposure by limiting the amount of collected data. This approach can include deleting

unnecessary data, reducing the potential impact of data breaches, and minimizing privacy violations for customers.

2. Promoting Ethical Data Usage: Data minimization fosters ethical data practices, which are essential for building trust and credibility. It ensures that organizations handle personal information responsibly, respecting individuals' privacy rights and promoting honesty and integrity in data-related activities.

3. Reduced Legal Liability: Minimizing the amount of data collected can mitigate legal liabilities associated with data breaches and privacy violations. By holding less data, organizations decrease the likelihood of regulatory non-compliance and potential legal challenges.

4. Cost Savings: Data minimization leads to cost savings for organizations. Storing and processing large volumes of data can be expensive in terms of infrastructure, resources, and maintenance. By collecting and retaining only necessary data, organizations can reduce operational expenses related to data management.

Principle of Data Minimization:

Data minimization is a foundational principle embedded within the General Data Protection Regulation (GDPR), essential for adhering to its core tenets of transparency and purpose limitation. GDPR mandates that organizations precisely define the scope of data collection and its intended uses (Sloot, 2013).

This principle necessitates that organizations collect only the minimum amount of personal data required to achieve specific objectives. Whether through transparent privacy policies accessible to users or cookie banners disclosing data collection practices, organizations are obligated to inform individuals about the types of data being collected and the purposes for which it is used. This approach underscores the importance of ethical conduct and responsible management of personal information to uphold stringent privacy standards (Qu et al., 2019).

4. **Anonymization and Pseudonymization**

Pseudonymization involves processing personal data in a way that renders it unidentifiable to a specific individual without additional information (GDPR, 2024a). This typically involves replacing directly identifying details like names, Social Security numbers, or dates of birth

with random codes. However, the data can still indirectly relate to the individual, maintaining its classification as personal data (Article 29 Working Party (Art. 29 WP), 2014).

Moreover, any supplementary information that could potentially identify the original person, such as an encryption key, must be stored separately from the data using technical or organizational safeguards (Tosoni, 2020). This segregation can occur even within the same organization to prevent accidental or unauthorized linkage of the data to an individual. A typical application of pseudonymization is key-coding data, commonly employed in fields like medical research, where personal information is labelled with codes. Various methods exist for pseudonymizing data, including encryption with a secret key, employing hash functions, or using keyed-hash functions with securely stored keys (Tosoni, 2020).

According to Recital 26 of the GDPR, data is considered anonymous if there is a reasonable certainty that it cannot be linked to an identified or identifiable natural person (GDPR, 2016). In data protection terms, data that are not capable of identifying individuals are typically termed ‘anonymous data’, and the process of transforming personal data into non-identifiable form is known as ‘anonymization’. Non-personal data fall outside the scope of the GDPR's application.

Anonymization involves aggregating information to the extent that specific details (like an individual's travel patterns) can no longer be linked to any particular person (European Medicines Agency, 2024). This type of data is frequently used in statistics or sales analyses to understand trends such as product preferences. Examples of anonymous data include insights into high-frequency trading in finance and data on precision farming, which optimizes the use of resources like pesticides and nutrients. When assessing whether an individual can still be identified from their data, all potential means of identification must be considered, whether by the data controller or others. Factors such as the time and technology needed for identification should be evaluated to determine the likelihood of re-identification (GDPR, 2016).

Pseudonymisation and anonymisation are often confused but play distinct roles under the GDPR. Pseudonymisation makes personal data unidentifiable unless additional information is available, while anonymisation ensures personal data cannot be re-identified. Understanding this difference is critical as it determines whether GDPR regulations apply to data processing. Misinterpreting pseudonymisation as anonymisation can lead to privacy risks if data is not adequately protected, potentially exposing personal information in

breaches. Conversely, incorrectly treating pseudonymised data as anonymised may restrict its use unnecessarily, affecting its value for research and other purposes. Table 1 shows the key difference between Pseudonymisation and anonymisation.

Table 1. Key difference between Pseudonymisation and anonymisation.

Pseudonymised Data	Anonymised Data
A patient's medical record with their name replaced by a unique identifier	Aggregated medical information with no personal identifying information, such as age and gender statistics of a patient population
A customer's transaction history with their name and address replaced by a unique identifier	Aggregated purchase data with no personal identifying information, such as the total amount spent by all customers in a specific geographic location
A person's driving record with their name and license number replaced by a unique identifier	Aggregated data on the average speed of vehicles on a specific road, without any personal identifying information
An employee's salary record with their name replaced by an employee number	Aggregated data on the average salary of employees in a specific department, without any personal identifying information
An individual's voting record with their name and address replaced by a unique identifier	Aggregated data on the distribution of political affiliations in a specific geographic location, without any personal identifying information

Source: (ISO, 2023).

5. **Transparency and Notice**

Policymakers and consumer advocates globally are increasingly advocating for greater transparency in service providers' data processing practices due to the economic importance of personal data and its implications for users (Executive Office of the President, 2015). They emphasize the need for regulations that empower individuals to actively control how their personal data is used and disclosed in the future (Podesta et al., 2015).

In Europe, the recent EU GDPR represents a pivotal regulation mandating enhanced transparency from service providers regarding their handling of personal data compared to previous practices (European Union, 2016; Robinson et al., 2009). The goal of transparency is to enable individuals to make more informed and privacy-conscious decisions regarding the sharing of their personal information (Tsai et al., 2010).

Transparency Enhancing Tools (TETs):

In 2012, Pulled proposed two privacy-preserving technologies (TETs): (a) a cryptographic system designed to facilitate transparency logging, and (b) a cryptographic scheme aimed at storing data for the Data Track tool within a cloud provider. The objective of the transparency

logging TET is to ensure that data controllers' processing activities are transparent to the individuals whose data is being processed (Pulls, 2012).

(a) The cryptographic system is a recent innovation aimed at achieving fully distributed transparency logging of data processing while safeguarding user privacy. It relies on formally verified cryptographic primitives and introduces a novel technique known as cascading. Cascading involves linking multiple cryptographic operations in a sequence, enhancing the security and privacy of data by adding layers of encryption or other cryptographic transformations. This approach not only strengthens the protection of sensitive information but also ensures that transparency logging remains resilient against potential privacy threats.

(b) The cryptographic scheme designed for the Data Track involves a specialized approach to managing data within a custom cloud storage environment. This scheme is structured to ensure a clear separation of responsibilities between writing to and reading from the storage. Specifically, the agents responsible for writing data to the storage are considered minimally trusted, implying they have limited access and authority. In contrast, the reader possesses full privileges for both reading and writing operations. The storage provider itself is viewed as a potential adversary, although it is assumed to be passive, meaning it behaves honestly but retains curiosity about the stored data. To address these security concerns, the scheme employs various formally verified cryptographic primitives. These components are utilized to achieve anonymous storage, which ensures that the stored data cannot be linked directly to specific individuals. Additionally, the scheme aims to hold the cloud provider accountable for maintaining data integrity, thereby enhancing overall security and trust in the storage system.

6. **Data Security**

Ensuring data security and preserving privacy are essential prerequisites for safeguarding end users in their business, economic activities, and everyday routines. The security and privacy requirements for data in both cloud computing and edge computing scenarios involve outsourcing user data to third parties like cloud or edge data centers, leading to concerns over ownership, control, and potential risks such as data loss, leakage, and unauthorized operations (e.g., replication, dissemination). These issues highlight

fundamental challenges in ensuring data confidentiality and integrity within edge computing environments (Bhatia & Verma, 2017; Mollah, 2018).

Data Security Steps in Edge Computing:

- Data Confidentiality: Ensures only authorized users and the data owner can access private information in edge computing. User data, stored on edge servers, faces risks like data loss and illegal operations due to separated ownership and control. To mitigate these risks, data confidentiality schemes such as encryption are used. Traditional methods include symmetric (AES, DES) and asymmetric (RSA, ECC) encryption, though they complicate data processing (J. Zhang et al., 2018). Recent advancements like identity-based (Boneh & Franklin, 2001), attribute-based (S. Wang et al., 2016), proxy re-encryption (Liang et al., 2015), and homomorphic encryption (A. Banerjee et al., 2017) provide secure data storage on untrusted edge servers.
- Data Integrity: Integrity guarantees the correct and consistent delivery of data to authorized users without any undetected modifications. Data integrity is critical in edge computing, where user data is transferred to edge servers and may be vulnerable to compromise (J. Zhang et al., 2018). It involves ensuring that data owners can verify the integrity and availability of outsourced data, preventing unauthorized modifications. In edge computing, research on data integrity focuses on four key aspects: batch auditing (Cong Wang et al., 2010), dynamic auditing(Q. Wang et al., 2011), privacy-preserving techniques (Yang & Jia, 2013), and solutions with low computational complexity (Lin et al., 2016).
- Secure Data Computation: Secure data computation is a critical concern in edge computing, where sensitive data from end users is often stored in encrypted form on edge servers. One major challenge is conducting secure keyword searches over encrypted data files without decryption. Researchers have developed searchable encryption methods, such as the secure ranked keyword search scheme (Cong Wang et al., 2012), which retrieve accurate search results based on specific criteria and indexing. Other advancements include attribute-based keyword search schemes (Changji Wang et al., 2013) for fine-grained data sharing, dynamic search methods (Kamara et al., 2012) for updating data dynamically, and proxy re-encryption with keyword search (Shao et al., 2010) to manage search privileges effectively.

- Authentication: Authentication verifies user identities. Authentication in edge computing, a complex distributed system with multiple trust domains, is crucial to prevent unauthorized access and erase malicious traces. Without robust authentication mechanisms, external adversaries can compromise sensitive resources, while internal adversaries can cover their tracks using legitimate access. Authentication methods such as single-domain (M Sravanthi & Bhaskar, 2016), cross-domain (Touceda et al., 2015), and handover (Vallent et al., 2021) authentication are essential to validate identities across different trust domains.
- Access Control: Access control regulates actions such as reading (confidentiality) and writing (integrity) within edge computing environments. Access control in edge computing faces significant security challenges due to its outsourcing nature. Without efficient authentication, malicious users can exploit edge or core infrastructure resources. Implementing fine-grained access control across multiple trust domains within distributed edge computing environments is crucial. Traditional mechanisms often limited to single trust domains may not suffice. Cryptography-based solutions such as attribute-based encryption (Yu et al., 2010), role-based encryption (Zhou et al., 2013), and TPM-based access control (Liqun Chen & Urian, 2015) provide flexible and granular access control in edge computing ecosystems.
- Privacy Requirement: Privacy requirements focus on preserving user information secrecy against honest but curious adversaries through security mechanisms like encryption, integrity auditing, authentication, and access control. Privacy preservation is a critical challenge in edge computing, exacerbated by the migration of sensitive data and personal information from edge devices to remote servers. Unlike other computing paradigms, edge computing faces additional risks due to numerous honest but curious adversaries, including edge data centers, infrastructure providers, service providers, and potentially malicious users. These authorized entities may exploit their access to gather sensitive information for various self-serving purposes, posing a threat to user privacy within an open ecosystem spanning different trust domains (J. Zhang et al., 2018). To mitigate these risks, privacy-preserving techniques in edge computing focus on several strategies. Data aggregation using homomorphic encryption enables analysis without requiring decryption, ensuring privacy during data processing.

Lightweight methods like probabilistic public key encryption (Bahrami & Singhal, 2015) and pseudo-random permutation (Olakanmi & Dada, 2019) are designed to protect sensitive information efficiently. In dynamic distributed environments, protecting user identities during authentication and management processes is crucial (Khalil et al., 2014; Park et al., 2013). Similarly, safeguarding location privacy is vital, given that user movements can be predicted based on frequent interactions with specific edge servers. These measures aim to uphold privacy amid the complex and interconnected nature of edge computing environments (Mollah et al., 2017).

Data Security Technologies- Cryptography:

This section explores five key cryptosystems: identity-based encryption (IBE), attribute-based encryption (ABE), proxy re-encryption (PRE), homomorphic encryption, and searchable encryption. These systems are essential for ensuring data confidentiality in cloud, edge, and distributed computing environments. Searchable encryption, a cryptography-based technique, is particularly notable for enabling secure data computation. IBE simplifies certificate management by using user identities as public keys, allowing secure communication without the need for key exchanges (Shamir, 1984). In ABE, access to encrypted data is controlled based on user attributes, supporting complex decryption policies to ensure only authorized users can access the data (Sahai & Waters, 2005).

Proxy re-encryption (PRE) allows ciphertexts encrypted under one key to be transformed into ciphertexts under another key using a semi-trusted proxy, facilitating secure data sharing without revealing the plaintext (Blaze et al., 2006a). Homomorphic encryption enables computations on encrypted data, ensuring secure data processing without decryption. This technique supports various operations like addition and multiplication on ciphertexts, preserving data privacy (Rivest et al., 1978). Lastly, searchable encryption allows querying encrypted data without compromising confidentiality, ensuring efficient and secure data retrieval in environments where data privacy is paramount (Song et al., 2000).

7. **Accountability**

Privacy-preserving computation with accountability enables processing private data without compromising its privacy, while also providing an audit trail. This trail allows third parties to verify the success of the computation and identify any bad actors who attempted to cheat (DIREC, 2023).

Accountability involves identifying a party with indisputable evidence to assume responsibility for misconduct. However, privacy requirements to protect users' personally identifiable information (such as name, email, and phone number) pose a threat to accountability. Users must maintain control over this information, deciding when, to whom, and where to disclose it. Anonymous user authentication complicates monitoring and tracking data usage, potentially encouraging misuse. Ghorbel et al. (2022) resolves the conflict between accountability and user privacy by introducing a trusted Auditing Party (AP). The AP facilitates recovering the identities of involved entities and revoking their attributes as needed.

The accountability phase includes two procedures: data tracking and attribute revocation. The BC-ABAC framework guarantees transparent accountability while maintaining user anonymity. It preserves user information privacy through anonymous authentication (Property E) and allows the AP to uncover a user's identity when necessary. The blockchain network ensures transparent accountability by providing a tamper-proof ledger that records data access and usage. Additionally, it enables collaboration among involved parties, such as service administrators, in the auditing process (Ghorbel et al., 2022).

An accountable privacy-preserving scheme for public information sharing systems:

Imine et al., (2020) proposed a secure, accountable privacy-preserving scheme using secret sharing and randomization techniques for anonymous and accountable public information sharing. After a one-time registration with the authority, entities can share information via externalization servers without needing the registration authority or third parties. Entities sign information with anonymous tokens, enabling servers to verify authenticity without compromising privacy. If anomalies are detected, the authority can trace entities despite anonymity. The solution requires minimal storage and communication overhead, and servers do not need to store user pseudonyms or temporary certificates. The scheme efficiently handles impersonation attempts and outperforms existing solutions (Imine et al., 2020).

8. **International Data Transfers**

A data transfer involves moving data across international borders and legal jurisdictions. This happens when an organization collects, stores, and processes data through third parties in different countries. Under the General Data Protection Regulation (GDPR), a transfer occurs

when an organization subject to GDPR shares personal data with another organization based in a non-EU country (Kaminska, 2024).

Data transfers are crucial for global trade, enabling the free flow of data across countries. However, these transfers must comply with personal data protection laws applicable in both the source and destination countries. Organizations must implement measures to protect data security and privacy in line with these laws. The GDPR allows data transfers between countries with adequate protection laws, as recognized by the European Commission. In countries without such recognition, organizations must use specific data transfer mechanisms to ensure compliance (European Union, 2024).

Data transfer involves moving data across international borders, occurring when organizations collect, store, and process data through third parties in other countries. Under the GDPR, a transfer happens when a GDPR-regulated organization shares personal data with an entity in a non-EU country. Such transfers are essential for global trade, enabling data flow across nations. However, they must comply with data protection laws in both the originating and receiving countries. Organizations must ensure data security and privacy according to these laws. The GDPR permits data transfers to countries with adequate protection laws recognized by the European Commission. In countries lacking this recognition, organizations must use specific data transfer mechanisms to remain compliant (OneTrust, 2022).

- Adequacy Decisions: The European Commission assesses whether non-EU countries offer GDPR-equivalent data protection, allowing data transfers without additional safeguards.
- Standard Contractual Clauses (SCCs): SCCs are agreements between data exporters and importers to ensure strict data protection, reviewed after the Schrems II ruling.
- Binding Corporate Rules (BCRs): BCRs are legally binding rules within multinational companies to protect personal data during internal transfers under the GDPR.
- Derogations: Derogations allow data transfers in specific cases, such as when explicit consent is given, but are not suitable for repeated or large-scale transfers.
- OneTrust Solutions: OneTrust Schrems II Solutions help organizations ensure international data transfers comply with the latest European Commission guidelines, offering tools and templates for necessary safeguards.

Under the DPDP Act in India, personal data can be transferred internationally for processing, except to countries specifically blacklisted by the Government of India. These restrictions

aim to protect personal data from being transferred to nations with inadequate data protection standards. Data transfers must be conducted under valid contracts that include provisions for indemnity in case of third-party breaches and specify secure methods of transfer. Additionally, the DPDP Act imposes obligations on Data Processors, such as implementing security safeguards, reporting data breaches to Data Fiduciaries, deleting data upon request, and adhering to India's blacklist of countries for data transfers (DLA Piper, 2024).

9. **Data Breach Notification**

A data breach occurs when there is unauthorized access to or loss of data containing sensitive personal information, potentially compromising the confidentiality or integrity of the data. California passed the first state data security breach notification law in 2002. Subsequently, many jurisdictions enacted similar laws, leading to the disclosure of over 2,676 data breaches and computer intrusions by various entities such as data brokers, businesses, retailers, educational institutions, government agencies, healthcare providers, financial institutions, nonprofits, utility companies, and Internet firms (Stevens, 2012).

A personal data breach under GDPR refers to a security incident where personal data is accidentally or unlawfully destroyed, lost, altered, disclosed, or accessed without authorization. Notification of a personal data breach involves the obligation to inform relevant authorities when data controllers or processors become aware of such incidents (Karyda & Mitrou, 2016).

The obligation to notify personal data breaches under the GDPR is clear and stringent. Data controllers must promptly inform the relevant supervisory authority as soon as they become aware of a breach, ensuring notification occurs within 72 hours unless there are valid reasons for a delay. This requirement applies when the breach is likely to result in risks to the rights and freedoms of individuals, extending beyond the GDPR's jurisdiction if necessary. Notifications must be comprehensive, detailing the breach's nature, contact information for key personnel like the data protection officer (DPO), types of data affected, the number of individuals affected, actions taken in response to the breach, and any other pertinent information to ensure transparency and accountability in data protection practices (i-SCOOP, 2024; Karyda & Mitrou, 2016).

10. **Legal Compliance**

Privacy-Preserving Data Sharing and Analysis (PPDSA) technologies facilitate the processing of datasets protected by diverse privacy laws, posing challenges for compliance across multiple legal frameworks. To mitigate this complexity, regulatory bodies should collaborate with developers, privacy experts, and users of PPDSA technologies to establish a shared understanding of their operation within existing regulations. Furthermore, adoption of these technologies necessitates heightened awareness among policymakers and legal professionals. Agencies, professional bodies, and legislatures should enhance their technical proficiency to aid in drafting laws, regulations, and assessment criteria. Emphasizing education and training will support the effective application of current regulations and inform future privacy legislation based on technological advancements (National Science and Technology Council, 2023).

An example of a Legal Compliant Ontology to Preserve Privacy for the Internet of Things: The emergence of Internet of Things (IoT) presents opportunities to enhance user efficiency and quality of life. However, detailed data analysis from IoT resources raises significant privacy risks. To empower data owners with control over their data, semantic modelling has been identified as a viable solution. Yet, many data owners lack expertise in privacy, which hampers their ability to assert their legal rights effectively. Semantic modelling becomes crucial for deducing necessary privacy protections.

Addressing these challenges, Loukil et al., (2018) introduced LIoPY, a privacy ontology designed to establish a standardized privacy vocabulary. LIoPY leverages description logic for standard reasoning technologies to ensure compliance with privacy requirements across the lifecycle of data collected by IoT resources. Initially applied in healthcare, LIoPY's applicability extends to various IoT contexts, providing a legal compliant framework to safeguard privacy (Loukil et al., 2018).

5. Ethical Foundations of Privacy: Balancing Principles in a Technological Era

Beyond legal compliance, understanding privacy-preserving principles reflects ethical considerations about how personal information should be treated. It demonstrates a commitment to respecting individuals' rights and expectations regarding their privacy. Ethics in privacy-preserving practices focus on ensuring that individuals' personal data is handled with respect, fairness, and integrity.

The application of privacy ethics, encompassing both those affected (moral patients) and those responsible (moral agents), depends on continuously refining and balancing ethical principles. Rapid technological advancements make fostering ethical awareness challenging, as constant change complicates establishing consistent practices. Consequently, ethical decision-making in privacy design will require ongoing attention to both knowledge and ethical considerations.

If we view privacy ethics as centered around ethical principles and epistemic contexts, its network encompasses various specified ethical issues. The principlist framework is significant because it provides a common normative basis across disciplines, professions, and governance bodies involved in privacy discussions. Balancing and specifying principles enable multiple perspectives on ethical priorities, such as non-maleficence, in any context. This ongoing ethical discourse supports collaborative decision-making (Nietzsche, 2021).

Reframing Privacy Ethics: Emerging Ethical Challenges:

Recognizing the expanded social and technological contexts, privacy concerns have shifted from focusing solely on individuals to encompassing interdependent networks and the impact of artificial intelligence. This shift necessitates a reconceptualization of privacy ethics based on universal human rights, as outlined in the Universal Declaration of Human Rights. These rights, which are legally enforceable, include respect for human dignity, individual freedom, justice, equality, and the right to participate in governance (United Nations, 2024).

In this section, we address specific challenges posed by twenty-first-century technologies, using an ethical framework organized around five principles: autonomy, justice, non-maleficence, beneficence, and explicability. Traditionally associated privacy concepts, such as anonymity, confidentiality, consent, and minimization, are now expanded to encompass broader ethical concerns. Specifically, this section argues for a shift in privacy ethics:

- From individual data protection to addressing multifaceted surveillance that violates human dignity.
- From individual consent to ensuring human agency and autonomy.
- From individual due process to ensuring social fairness and non-discrimination.
- From individual risk assessments to protecting vulnerable populations with non-maleficent goals.

- From singular contexts of intrusion to collective responsibilities for environmental, social, and cultural well-being aligned with beneficent goals.
- From limited scope and purpose to ensuring data integrity, provenance, and accountability in algorithmic processes.

The following section will delve into the five ethical principles: autonomy, justice, non-maleficence, beneficence, and explicability.

1. **Respect for Autonomy in Survey Design: Ensuring Informed and Willing Participation**

To encourage trust and accurate responses from survey participants, statistical agencies should prioritize respecting privacy. This involves minimizing the time and effort required to respond and providing clear information so participants can make informed decisions about participating. Agencies should clarify whether participation is mandatory or voluntary, how data will be used, who will access it, and the expected duration and frequency of interviews (Ali & Bhaskar, 2016). They should design surveys with clear, non-intrusive questions that fit respondents' understanding, offer alternative response modes like Internet or smartphone, and utilize existing data sources where possible to reduce respondent burden. Using well-trained staff and applying principles from cognitive psychology to questionnaire design can further enhance accuracy and respondent cooperation (National Research Council, 1984).

Despite privacy concerns impacting survey response rates, agencies must continue to uphold confidentiality pledges and comply with laws like the E-Government Act and Federal Cybersecurity Enhancement Act, which safeguard personal information from unauthorized access and misuse. Effective communication of these policies to respondents remains a crucial challenge (National Research Council, 2008).

Protecting and respecting the autonomy of human research participants:

Collecting data from individuals for federally funded research is governed by regulations and best practices developed over more than 50 years, primarily outlined in the "Common Rule" (45 CFR 46) (National Research Council, 2023). Revised in January 2017, these regulations mandate protecting participants' privacy, maintaining data confidentiality, minimizing risks, ensuring equitable participant selection, and obtaining informed consent. Most federally

funded human research must be reviewed by an independent institutional review board (IRB) to ensure ethical compliance.

Even if not all data collections require IRB review, federal statistical agencies should adhere to the principles of the Common Rule. For activities needing IRB approval, agencies should collaborate with the IRB to apply regulations without discouraging participation. For instance, implied consent is often sufficient for general mail and telephone surveys, and providing advance information about the survey can help potential respondents decide whether to participate. Agencies should ensure that data collected are compiled and made accessible, and if data quality is insufficient, they should explain why the results are not published (U.S. Department of Health and Human Services, 2017).

Respecting the providers of alternative data:

To adopt a new paradigm using multiple data sources for federal statistics, statistical agencies must develop procedures that respect the constraints of data providers, including federal agencies and non-federal entities. Cooperation should not be assumed. Agencies holding useful records should collaborate continuously, starting with a memorandum of understanding, fostering mutual respect and trust to improve data quality for both parties (U.S. Office of Management and Budget, 2014).

Using administrative records requires considering whether informed consent from the original data providers is necessary. The U.S. Office of Management and Budget outlines scenarios where statistical use of records may qualify under the “routine use” exception of the Privacy Act, or may require a modified System of Records Notice. In some cases, obtaining new consent may be necessary (National Academies of Sciences, Engineering, 2017).

2. Autonomy as Dignity: From Data Protection to Multifaceted Forms of Intrusion

For the past 50 years, privacy concerns have focused on protecting electronic records in various sectors to ensure individual privacy rights. Today, these concerns are more complex due to diverse data types (numeric, text, voice, image, biometric) and advanced collection technologies (Kak, 2020). Modern society is now surrounded by video cameras, facial recognition systems, RFID chips, electronic toll collectors, and smartphones with location tracking, leading to extensive surveillance in both public and private spaces. Privacy ethics must now address these multifaceted surveillance methods as intrusions on personal dignity

and autonomy, encompassing concerns about one's person, identity, and information (Polyakova & Meserole, 2019).

Facial recognition technologies (FRTs) exemplify these challenges. They use biometric data to identify individuals by matching images against stored data. This raises significant privacy concerns because biometric data, such as fingerprints and facial images, are uniquely identifiable, sensitive, and difficult to secure. Misuse of this data can lead to identity theft and other issues. FRTs used by governments for tracking and surveillance can create power imbalances and act as tools of social control, leading to digital authoritarianism.

For instance, in China, FRTs are part of a social scoring system that monitors citizen behaviour and imposes consequences for non-compliance (Kostka et al., 2021). In India, the Aadhaar system collects extensive biometric information from citizens, raising concerns about government overreach (S. Banerjee & Sharma, 2019). FRTs often operate continuously and without consent, increasing the risk of privacy intrusions. Additionally, issues with accuracy and reliability, particularly for women and people of colour, have led to calls for moratoriums on FRT use in public spaces. The use of such technologies can result in structural violence, loss of freedom, and increased surveillance, posing significant ethical and privacy risks (Ranganathan, 2019).

3. Autonomy as Agency: From Consent to Access

A prominent privacy concern involves the expectation that individuals are informed and give consent when their personal information is accessed. Informed consent assumes individuals can decide when and how much information to share, a principle foundational to research ethics and industry terms of service (Westin, 1968). However, today's widespread and often invisible data collection makes meaningful consent difficult. Withholding consent can deny access to essential services, and complex terms of service often obscure data-sharing permissions, rendering consent ineffective in protecting individual privacy rights (Nissenbaum, 2018).

Respecting autonomy now requires shifting focus from consent to access, recognizing that control over information is crucial. Key questions include the role individuals play in data use decisions, the level of control humans retain in automated systems, and how systems adjust to potential privacy intrusions (Fiesler et al., 2020; Friedman & Nissenbaum, 1996).

The ethical concern lies in balancing autonomy between consent and agency. Individuals need the capability to act on their decisions and the agency to intervene with automated systems making determinations about them (IEEE, 2017). For instance, self-driving vehicles must be designed to reflect community values, ensuring trustworthiness and human oversight. An iterative design process assessing ethical implications and following aligned standards is crucial. Human-centric computing emphasizes designing technologies with consideration of human impact, centering human values in the design process through approaches like value-sensitive design and privacy by design (OIPC, 2014; Shneiderman, 2020).

4. Justice: From Material Risk to Fairness and Due Process

Due to increasing evidence and concerns about unfairness in technologies and algorithms, recent discussions have called for reorienting and broadening ethics to include justice in social, racial, economic, and environmental (Ada Lovelace Institute, 2020; Hao, 2021). Some term this as information (Butcher, 2009) or algorithmic justice (Re et al., 2019). These discussions focus on technical methods to address fairness, bias, and discrimination in algorithms and the negative impacts on individuals and groups from flawed systems (Algorithmic Justice League, 2024).

For instance, Buolamwini and Gebru's research showed that a facial recognition system had a 35% error rate for darker-skinned females, compared to 1% for lighter-skinned males, indicating gender and racial biases in these technologies (Buolamwini & Gebru, 2018). Similarly, a predictive algorithm used in U.S. courts for recidivism rates was more likely to incorrectly label Black defendants as high-risk compared to White defendants, highlighting societal risks from biased algorithmic decisions (Angwin et al., 2016).

Algorithmic bias also affects non-marginalized populations. For example, Amazon's hiring algorithm unintentionally widened gender gaps by favouring male applicants due to the training dataset's male-dominant makeup. This issue stemmed from missing data rather than inaccuracies (Hamilton & Sodeman, 2020). Algorithmic bias can also violate due process. In states such as California, New York, and Texas, automated teacher evaluations led to terminations without informing employees or allowing for accountability, raising ethical concerns about opaque systems (F Pasquale, 2016; Reisman et al., 2018).

On a societal level, China's social credit scoring system (SCS) illustrates large-scale privacy and human rights violations (Rachel, 2017). The government assigns "trustworthiness" scores based on extensive data, impacting citizens' freedom of movement, employment, education, and reputation. This underscores the high stakes of state surveillance and the need for ethical scrutiny in algorithmic applications (Y. Chen & Cheung, 2017; Pasquale, 2020).

5. Non-maleficence and Beneficence: From Individual Risk to Collective Societal Good

Non-maleficence:

Privacy ethics have traditionally focused on assessing risks to individuals and ensuring the safety, robustness, and protection of vulnerable populations. Much of the legal discourse around privacy protection and rights revolves around the harmful consequences experienced when privacy is breached. However, the concept of harm is often narrowly defined, leading to unpunished violations. By expanding the ethical focus to include non-maleficence which emphasizes preventing potential harms to society at large in research and technological designs we can address broader concerns beyond individual risk assessments and consider long-term social, intellectual, and political impacts (Polyakova & Meserole, 2019). Modern technologies amplify privacy concerns, particularly for young, vulnerable, and marginalized populations. For instance, authoritarian regimes use surveillance technologies to identify and target individuals, as seen with the Uyghur population in China, who face severe persecution through extensive government surveillance (Mozur, 2019).

Significant privacy concerns arise at the intersection of humans and technologies, especially for the young, vulnerable, and marginalized populations. Contemporary technologies exacerbate these concerns, particularly in authoritarian regimes where surveillance tools can have dangerous consequences. For example, facial recognition and other surveillance technologies have been used in China to identify, persecute, and imprison Uyghurs, who are considered enemies of the Communist Party (Roberts, 2021). This is facilitated by extensive government surveillance, including CCTV cameras with facial recognition in public spaces and spyware on smartphones, which tracks location, communication, and media use (Floridi et al., 2018).

Beneficence:

If non-maleficence requires moral agents to simply avoid harm, beneficence emphasizes actively doing good. Beneficence involves balancing individual and collective concerns to design research that benefits society's well-being. This approach goes beyond avoiding risks to enhancing social, environmental, and cultural well-being. In research, this means not just asking how to avoid risks but also how to conduct work that generates social good. In industry, movements such as ICT4All and AI4Good promote designing technologies for social and economic development, especially for underserved populations. Fields including human-computer interaction (HCI) call for culturally sensitive, accessible, and environmentally friendly policies (Floridi et al., 2018).

Non-maleficence and beneficence intersect in privacy practices, balancing avoiding harm with developing protective measures. During the COVID-19 pandemic, technologies like contact tracing and symptom tracking aimed to manage public health risks but raised privacy concerns (T. C. Li, 2021). These health surveillance systems needed to balance short-term data sharing against long-term risks of permanent surveillance. The history of surveillance after crises, such as post-9/11 measures, highlights the ethical tension in balancing public health benefits with privacy risks (Boudreaux et al., 2020).

6. Explicability: From Data Transparency to Process Intelligibility

Ethical values are closely linked to epistemic values, which concern what and how we know. Privacy ethics have traditionally emphasized transparency (notifying individuals about data collection, use, and dissemination) and accountability (ensuring compliance with privacy protection). In the modern era, where complex systems are often opaque, these principles should also encompass intelligibility (understanding how systems work) and clear data provenance (identifying who is responsible) (U.S. Department of Health, 1973).

This collective principle is termed explicability, which ensures individuals understand the decision-making processes that significantly impact their lives, such as employment, housing, credit, and criminal sentencing. Explicability supports other ethical principles by requiring transparency in algorithmic decision-making, enabling accountability for unjust outcomes, and ensuring technologies benefit society while avoiding harm (Floridi et al., 2018; Floridi & Cowls, 2019).

For instance, the COMPAS algorithm, used in criminal justice to predict recidivism, has been shown to be no more accurate than human judgment and has produced racially biased outcomes. Individuals affected by such decisions often cannot obtain explanations due to claims of proprietary information, violating their right to due process (Dressel & Farid, 2018). The factual and epistemic contexts of privacy issues are crucial in applying ethical principles. Explicability in data is essential for both ethical reception and transmission, particularly when artificial agents are involved. As ethical challenges evolve, they add complexity to principles such as transparency and comprehensibility, affecting research policy and practice (Barocas, 2016).

6. Related work: Recent Advances in Privacy-Preserving

Several proposals in the literature have addressed privacy concerns in various contexts. Liu et al. (2012) proposed an anonymous payment system aimed at enhancing the location privacy of electric vehicles (Liu et al., 2012). De Fuentes et al. (2017) introduced PRACIS, focusing on privacy-preserving data forwarding and aggregation for cybersecurity information sharing (de Fuentes et al., 2017).

Nicanfar et al. (2013) developed a robust privacy-preserving authentication scheme for electric vehicles and power stations (Nicanfar et al., 2013). Rottondi et al. (2014) proposed a security infrastructure for privacy-friendly vehicle-to-grid interactions (Rottondi et al., 2014). Gope (2019) presented a privacy-preserving security architecture for device-to-device communication in fog computing (Gope, 2019).

In Wang et al. (2015), an accountable privacy-preserving communication scheme for smart grids was proposed, involving local aggregators (LAG), central aggregators (CAG), and electric vehicles using pseudonyms (H. Wang et al., 2015). Aslam and Zou (2009) suggested a distributed certificate architecture for VANETs, focusing on temporary pseudonyms for vehicles in specific areas (Aslam & Zou, 2009). Sucasas et al. (2016) proposed pseudonym-based privacy-preserving authentication for VANETs, addressing pseudonym generation frequency issues (Sucasas et al., 2016). Guan et al. (2019) developed a device-oriented anonymous privacy scheme for IoT systems (Guan et al., 2019). Salem et al.

(2010) proposed a non-interactive authentication scheme for V2V networks, emphasizing dynamic key management for group privacy (Salem et al., 2010).

He et al. (2016) proposed an accountable, privacy-preserving authentication framework for wireless access networks, suitable for group-based architectures (He et al., 2016). Shen et al. (2018) proposed an anonymous and accountable group data sharing scheme in the cloud, utilizing group signatures for anonymity and accountability, albeit requiring interaction with a group manager for data sharing events (Shen et al., 2018).

7. ISO/IEC Standard: Enhance Data Protection

Definition and Benefits:

ISO/IEC standards are developed by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). These standards offer specifications, guidelines, and best practices for a variety of products, services, and processes. They aim to ensure that products and services are safe, reliable, high-quality, and compatible with each other. Furthermore, these standards help maintain consistency and quality worldwide, regardless of the production or usage location. ISO/IEC standards also guide organizations in complying with national and international laws, regulations, and guidelines (6clicks, 2024b).

The ISO standards in the 27000 series form essential frameworks for information security and privacy management in organizations. ISO 27001 establishes an ISMS for securing information assets, while ISO 27002 provides detailed controls for implementation. ISO 27005 guides risk assessment and treatment, and ISO/IEC 27017 offers cloud-specific security controls. ISO/IEC 27701 extends privacy management controls, ISO/IEC 27018 focuses on cloud privacy, and ISO/IEC 29100 defines terms for PII management. ISO/IEC 20889 outlines de-identification techniques, and ISO/IEC 27555 provides guidelines for secure data deletion. These standards collectively enhance data protection and compliance efforts, crucial for organizations worldwide. The following section outlines examples of ISO/IEC standards categorized by topic area, providing an overview, emphasizing their importance, outlining principles, and describing their functions.

Cybersecurity Standards:

i. **ISO/IEC 27001:**

- Manages information security programs for organizations of any size and complexity. Organizations can achieve certification for compliance. It works with ISO/IEC 27002 for security controls.
- While it primarily focuses on ensuring the confidentiality, integrity, and availability of information assets, the implementation of security controls prescribed by ISO 27001 can indirectly contribute to privacy preservation by safeguarding personal data against unauthorized access, disclosure, or destruction.

ISO/IEC 27001 Overview:

Released in October 2022, ISO/IEC 27001 is the primary international standard for information security, co-published by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). Part of the ISO/IEC 27000 series, it provides comprehensive guidelines for managing all aspects of security, titled "ISO/IEC 27001 – Information security, cybersecurity, and privacy protection - Information security management systems - Requirements."

The ISO 27001 framework sets the standards for creating, implementing, operating, and improving an Information Security Management System (ISMS), aiming to protect information systematically and cost-effectively across any industry (Kosutic, 2024).

Importance of ISO 27001:

ISO 27001 helps companies protect valuable information and offers certification to prove data security to customers and partners. Individuals can also get certified, demonstrating their ISMS skills. Being an international standard, ISO 27001 is globally recognized, enhancing business opportunities (Kosutic, 2024).

Principles of ISO 27001:

The standard focuses on securing people, processes, and technology through the C-I-A triad:

1. Confidentiality: Protects data and systems from unauthorized access using controls like multifactor authentication, security tokens, and data encryption.
   - *Example Risk:* Criminals obtaining and selling client login details on the Darknet.

2. Integrity: Ensures data accuracy, trustworthiness, and completeness by preventing errors and manipulation, allowing access only to authorized personnel.
    - *Example Risk:* A staff member accidentally deleting a row in a file or database.
3. Availability: Maintains and monitors ISMS, removes security process bottlenecks, updates firmware, ensures business continuity with redundancy, and minimizes data loss with backups and disaster recovery.
    - *Example Risk:* An enterprise database going offline due to server issues and insufficient backups.

A compliant ISMS preserves confidentiality, integrity, and availability by applying a risk management process, assuring that risks are effectively managed (Hanna, 2024; Kosutic, 2024).

Function of ISO 27001:

ISO 27001 ensures the confidentiality, integrity, and availability of information assets through a systematic risk assessment and implementation of appropriate security controls. This foundational principle of risk management helps organizations identify, evaluate, and address risks to enhance their information security posture (Hanna, 2024).

ii. **ISO/IEC 27002:**

- A catalog of information security controls providing high-level guidance. Often used with ISO/IEC 27001 but can be a primary guideline on its own.
- These controls include encryption, access control, and incident management, which are essential for protecting personal data and ensuring privacy.

ISO/IEC 27002 Overview:

ISO/IEC 27002:2022 is an information security standard by the ISO and IEC, closely associated with ISO 27001. It provides guidance on implementing an ISO 27001 ISMS with a reference set of controls for information security, cybersecurity, and privacy protection, based on best practices. Though ISO 27002 is not certifiable on its own, complying with its guidelines helps organizations meet ISO 27001 certification requirements (Harvey, 2024).

Importance of ISO/IEC 27002:

Organizations handling data face inherent security risks. Implementing an Information Security Management System (ISMS) is essential to protect information confidentiality, availability, and integrity. However, newcomers to information security often find the breadth of ISMS implementation daunting. Managing an ISMS encompasses many aspects, which can leave managers uncertain about where to start. A practical way to begin is by adopting the controls outlined in ISO/IEC 27002. This standard offers widely accepted best practices for information security, cybersecurity, and privacy management (Harvey, 2024; Kirvan, 2024).

Principles of ISO/IEC 27002:

The ISO developed the 27002-standard based on the information security triad, which includes three foundational principles: confidentiality, data integrity, and availability (DataGuard, 2024).

1. Confidentiality ensures that only authorized individuals can access data, often through legal agreements, professional standards like doctor-patient confidentiality, or technological safeguards such as encryption and password protection.
2. Data integrity ensures that data remains accurate and complete over time, with measures in place to prevent the entry of erroneous information.
3. Availability ensures that authorized users can access data when needed, such as ensuring patients can access their health data within 30 days of request.

Function of ISO/IEC 27002:

ISO 27002 serves as a framework for implementing the controls specified in ISO 27001. While ISO 27001 focuses on establishing information security management systems (ISMS) and their objectives, ISO 27002 offers detailed guidance on selecting, implementing, and managing these controls based on best practices. Together, these standards provide organizations with a comprehensive foundation to enhance risk management, identify threats, and mitigate vulnerabilities effectively (Kirvan, 2024).

iii. **ISO/IEC 27005:**

- ISO/IEC 27005 provides guidance on managing information security risks based on ISO/IEC 27001 and ISO 31000.
- By conducting comprehensive risk assessments, organizations can identify vulnerabilities in data handling processes and implement appropriate controls to protect personal data.

ISO/IEC 27005 overview:

This document offers guidelines for managing information security risks, aligning with ISO/IEC 27001 principles. It aims to support effective implementation of information security through a risk management approach. Familiarity with concepts, models, processes, and terminology from ISO/IEC 27001 and ISO/IEC 27002 is crucial for comprehending this document fully. It applies to diverse organizations, including commercial enterprises, government agencies, and non-profits, all seeking to mitigate risks that could threaten their information security (IT Governance UK, 2024).

Importance of ISO/IEC 27005:

ISO/IEC 27005 is an international standard for information security risk management, offering guidance on implementing a risk management system within an organization. Based on ISO/IEC 27001, it provides a framework for assessing, managing, and responding to information security risks. The standard helps organizations apply ISO/IEC 27001 principles, select and implement controls, develop risk management policies, and monitor their effectiveness. It aims to ensure efficient and effective information security risk management, particularly for risks related to information technology and other information-related activities (6clicks, 2024a).

Principles of ISO/IEC 27005:

While ISO 27005 does not prescribe a specific risk management methodology, it suggests an ongoing information risk management process encompassing six key components (ISO, 2024a):

1. Context establishment
2. Risk assessment
3. Risk treatment
4. Risk acceptance
5. Risk communication
6. Risk monitoring and review

Function of ISO/IEC 27005:

The framework can be divided into the following steps (SISA, 2024):

1. Risk analysis, which includes:
    - Risk identification

- Risk estimation

2. Risk evaluation

ISO/IEC 27002:2005, Code of practice for information security management, recommends examining the following during a risk assessment (SISA, 2024):

- Security policy
- Organization of information security
- Asset management
- Human resources security
- Physical and environmental security
- Communications and operations management
- Access control
- Information systems acquisition, development, and maintenance
- Information security incident management
- Business continuity management
- Regulatory compliance

iv. **ISO/IEC 27017:**

- Provides cloud security controls, building on ISO/IEC 27002. It can be used with ISO/IEC 27001, as an extension to ISO/IEC 27002, or standalone for cloud security.
- Cloud computing often involves the processing of large volumes of personal data, and adherence to ISO/IEC 27017 helps organizations implement measures to protect privacy in cloud services.

ISO/IEC 27017 overview:

ISO/IEC 27017 is a security framework for organizations using cloud services, providing consistent and comprehensive guidelines to keep customers safe. Derived from ISO/IEC 27002, it adds specific cloud security controls and is part of the ISO/IEC 27000 family of standards for best practices in information security management. Published by the ISO and IEC under the ISO/IEC JTC 1/SC 27 subcommittee, this standard guides both cloud service customers and providers in implementing these controls. It aligns security management across cloud computing, and both virtual and physical networks, ensuring robust safety measures and risk-based analysis for online security.

Importance of ISO/IEC 27017:

ISO/IEC 27017 is a globally recognized framework that reduces data breach risks and enhances customer trust by demonstrating a commitment to information security. It covers issues such as asset ownership, asset removal after contract termination, and virtual environment security. The framework specifies administrative operations for managing cloud environments and hardening virtual machines according to business needs. Both cloud service providers and users need to show they are minimizing data breach risks. Based on ISO 27001 and ISO 27002, ISO 27017 demonstrates best practices for cloud security but does not replace ISO/IEC 27002 requirements (Alliantist, 2024).

Principles of ISO/IEC 27017:

ISO/IEC 27017 provides organizations with a systematic and uniform method for enhancing customer security, with a strong emphasis on cloud and data security. It specifically applies to both cloud service providers and customers.

The framework comprehensively outlines the expectations that customers can have from their cloud service providers, as well as the responsibilities and duties customers must fulfil to establish and uphold a secure cloud environment (Wadhwa, 2024).

Function of ISO/IEC 27017:

ISO 27017 stresses the importance of organizations categorizing their data according to its sensitivity and applying suitable protective measures. This involves identifying and classifying data based on its criticality and sensitivity, and implementing access controls and encryption methods accordingly (6clicks, 2024c).

Privacy Standards:

v. **ISO/IEC 27701:**

- This standard is an extension to ISO/IEC 27001 and ISO/IEC 27002 and provides controls for managing personal information within the context of an organization. Organizations may be a “PII controller” and/or a “PII processor” (as defined by ISO/IEC 29100).
- It helps organizations establish a Privacy Information Management System (PIMS) to ensure compliance with privacy regulations (e.g., GDPR) and protect individuals' privacy rights.

ISO/IEC 27701:2019 overview:

ISO/IEC 27701 is an international standard specifically related to data privacy and preserving privacy. It provides comprehensive guidelines for establishing, implementing, maintaining,

and continually improving a Privacy Information Management System (PIMS). As an extension of ISO/IEC 27001 and ISO/IEC 27002, it integrates privacy management within an organization's information security management system. By offering guidance on protecting privacy and managing personal information, ISO/IEC 27701 helps organizations demonstrate compliance with global privacy regulations (NQA, 2021).

ISO/IEC 27701 is an international standard for privacy management systems, offering guidance on protecting privacy, managing personal information, and ensuring compliance with global privacy regulations.

Benefits of Implementing PIMS:

- Build and maintain customer trust by ensuring PII is used only for its intended purpose.
- Promote personal data management within the organization’s culture.
- Demonstrate compliance with GDPR and other data protection laws.
- Protect the confidentiality and integrity of PII.
- Mitigate PIMS security risks.
- Gain a competitive advantage by showcasing best practices (Bureau Veritas India, 2020).

Importance of PIMS:

With the exponential growth in personal information collection and data processing, privacy concerns have risen. Implementing a PIMS per ISO/IEC 27701 helps organizations assess, treat, and reduce privacy risks related to personal information (Bureau Veritas India, 2020).

Applicability:

ISO/IEC 27701 applies to all organizations, including public and private companies, government entities, and not-for-profits, that act as PII controllers or processors within an ISMS (Bureau Veritas India, 2020).

Compatibility:

ISO/IEC 27701 is compatible with other management system standards, enabling organizations to align or integrate their PIMS with other management system requirements (Bureau Veritas India, 2020).

Principles of ISO/IEC 27701:

The controls and principles defined in ISO 27701 are consistent with the principles outlined in current data protection laws globally, such as GDPR and CCPA (S. Mishra, 2023).

The main principles of ISO 27701 certification are as follows:

- Privacy Policy: Create and maintain a clear and thorough privacy policy that details the organization's commitment to safeguarding personal information and adhering to privacy laws.
- Scope Definition: Clearly outline the boundaries and responsibilities of the Privacy Information Management System (PIMS) to manage privacy effectively.
- Leadership and Commitment: Top management must demonstrate leadership and commitment by supporting privacy goals and allocating necessary resources.
- Legal and Regulatory Compliance: Ensure compliance with applicable privacy laws, regulations, and standards across all operational regions and stay updated on changes.
- Privacy Objectives and Targets: Set measurable objectives to enhance privacy performance and compliance, aligning them with the organization's privacy policy and significant privacy risks.
- Privacy Impact Assessment (PIA): Conduct PIAs to assess how data processing activities may impact individuals' privacy rights, using findings to implement appropriate safeguards.
- Privacy by Design and by Default: Integrate privacy considerations into product, service, and system design, ensuring privacy features are enabled by default.
- Operational Controls: Implement controls and procedures to manage and protect personal information throughout its lifecycle—from collection and processing to storage and sharing.
- Data Subject Rights: Establish mechanisms for individuals to exercise their privacy rights (e.g., access, rectification, erasure) and ensure timely responses to requests.
- Third-Party Management: Collaborate with third-party processors to ensure they meet privacy standards through contractual agreements and oversight.
- Documentation and Record Keeping: Maintain comprehensive documentation of the PIMS, including policies, procedures, privacy impact assessments, and records of privacy incidents.
- Training and Awareness: Provide training to employees and stakeholders on their roles in privacy protection and compliance with privacy laws.

- Incident Response and Notification: Develop procedures for responding to privacy incidents and breaches, including notifying authorities and affected individuals as required by law.

Advantages of ISO 27701 Certification and GDPR Compliance:

ISO 27701 certification offers significant advantages across industries.

a. Global Recognition: It is a globally respected standard for managing privacy information systems, showcasing leadership in the field.
b. Stringent Evaluation: Certification involves rigorous assessments that highlight the thoroughness of operations and adherence to high standards.
c. Flexibility: It aligns with both global standards and local jurisdiction requirements, ensuring compliance across different regulatory landscapes.

ISO 27701 is crucial for organizations handling Personally Identifiable Information (PII), providing a structured framework for data management and privacy protection. It enhances existing information security management systems to meet privacy requirements, supporting compliance efforts such as GDPR.

GDPR has reshaped data privacy norms, imposing substantial penalties for non-compliance, including fines up to €20 million or 4% of annual global turnover. Certification under ISO 27701, while not a guarantee of GDPR compliance, offers a robust system for managing data protection risks and demonstrates proactive measures to regulators and stakeholders.

Implementing a Privacy Information Management System (PIMS) under ISO 27701 empowers organizations to mitigate risks associated with personal data handling. Additionally, organizations may explore BS 10012:2017 with Annex A1:2018 as an alternative approach for standalone Privacy Information Management Systems outside the scope of ISO 27001.

vi. ISO/IEC 27018:

- ISO/IEC 27018 extends the privacy controls of ISO/IEC 27002 specifically for cloud services.
- It provides guidelines for cloud service providers (CSPs) on how to handle personal data in compliance with privacy principles.

- This standard helps CSPs implement measures to protect personal data from unauthorized access, processing, or disclosure.

ISO/IEC 27018 overview:

This document outlines control objectives and guidelines for protecting Personally Identifiable Information (PII) in public cloud computing, aligned with ISO/IEC 29100 privacy principles. It builds upon ISO/IEC 27002 and addresses regulatory requirements applicable to PII processors in various information security risk environments. It applies to organizations of all types and sizes that provide PII processing services via cloud computing, under contract. While relevant to PII controllers, it does not encompass additional PII protection obligations that may apply specifically to them (ISO, 2019).

Importance of ISO/IEC 27018:

ISO 27018 offers several key benefits in your compliance framework:

- Boosted Customer Confidence: Certification demonstrates your ability to securely manage Personally Identifiable Information (PII), enhancing trust and differentiation from competitors.
- Simplified Global Operations: Universal guidelines ease entry into the global marketplace and facilitate international contract negotiations.
- Streamlined Contract Processes: Conformance reduces the need for extensive inquiries about PII handling practices, speeding up client onboarding.
- Enhanced Cyber Insurance Accessibility: Certification enhances credibility with cyber insurance providers, potentially simplifying coverage processes for data breaches or privacy violations.
- Lower Barrier to Implementation: Compared to ISO 27701, ISO 27018 focuses specifically on PII controls in public cloud environments, making it a more accessible option for organizations not seeking full management system requirements (HICKS, 2024).

Principles of ISO/IEC 27018:

The core principles and requirements of ISO 27018 include:

- Establishing a framework for processing personal data in cloud services, covering privacy, security, data protection, and compliance.

- Implementing appropriate technical and organizational measures to protect personal data processed by cloud service providers.
- Providing customers with transparent information on the usage and storage of their personal data.
- Enabling customers to exercise control over their personal data in accordance with legal requirements.
- Mandating adequate remedies for customers in case of personal data breaches or misuse.
- Promoting transparency between providers and customers regarding the collection, use, and sharing of personal data.
- Requiring cloud service providers to document changes affecting customers' personal data.
- Setting up procedures for monitoring compliance with ISO 27018 requirements within organizations.
- Establishing an independent audit process to verify compliance with the standard (Scytale, 2024).

Function of ISO/IEC 27018:

ISO/IEC 27018:2019 is a set of guidelines designed to safeguard personally identifiable information (PII) in public cloud computing, aligning with ISO/IEC 29100 privacy principles. It outlines requirements for public cloud service providers (CSPs) to protect PII, especially concerning international PII transfers, and details technical measures for securing PII stored in the cloud (Exoscale, 2024).

vii. ISO/IEC 29100:

- o ISO/IEC 29100 establishes a framework for privacy and defines essential terms like "personally identifiable information (PII) controller" and "PII processor."

- o It provides guidelines for organizations on how to manage and protect PII in accordance with privacy principles, thereby ensuring privacy-preserving practices are integrated into organizational processes.

ISO/IEC 29100 overview:

ISO/IEC 29100:2011 establishes a privacy framework that accomplishes several key objectives. It standardizes privacy terminology, clarifies the roles of entities involved in

processing personally identifiable information (PII), outlines considerations for safeguarding privacy, and references established privacy principles in information technology. This standard is relevant to individuals and organizations engaged in various aspects of information and communication technology, including specification, procurement, architecture, design, development, testing, maintenance, administration, and operation of systems or services requiring privacy controls for PII processing (ISO, 2024d).

Importance of ISO/IEC 29100:

Benefits of ISO/IEC 29100 Certification to Your Organization:

- Enhances privacy protection across information and communication systems
- Safeguards Personally Identifiable Information (PII)
- Boosts organizational credibility
- Reduces security breaches
- Establishes additional standards for privacy
- Defines and categorizes privacy features
- Integrates privacy considerations into security guidelines
- Assures customers of robust privacy protections for information and communication systems

Principles of ISO/IEC 29100:

The standard outlines 11 privacy principles, covering aspects such as consent, purpose specification, data minimization, limitations on use and disclosure, data accuracy, transparency, individual rights, accountability, information security, and privacy compliance (Drozd, 2017).

Function of ISO/IEC 29100:

SO/IEC 29100:2011 establishes a privacy framework that:

- Standardizes privacy terminology.
- Defines the roles of participants involved in processing personally identifiable information (PII).
- Specifies considerations for safeguarding privacy.
- References established principles of privacy in information technology.

ISO/IEC 29100:2011 Edition 1 is applicable to individuals and organizations engaged in specifying, procuring, architecting, designing, developing, testing, maintaining, administering, and operating information and communication technology systems or services that necessitate privacy controls for PII processing. The reference document ISO/IEC 29100:2011/Amd 1:2018, spanning 21 pages, was last updated in 2018. ISO/IEC 29100 Edition 2 is currently under development as ISO/IEC DIS 29100 (Nolasco, 2023).

viii. **ISO/IEC 20889**

- This standard defines terms for de-identifying personal data as well as descriptions of techniques.
- De-identification is a critical privacy-preserving technique used to anonymize data so that individuals cannot be re-identified.
- This standard helps organizations implement effective de-identification measures to protect privacy while using or sharing data.

ISO/IEC 20889 overview:

This document outlines privacy-enhancing data de-identification techniques aligned with the principles of ISO/IEC 29100. It includes terminology, categorizes de-identification methods based on their attributes, and addresses their effectiveness in mitigating re-identification risks. Applicable to organizations of varying scales and sectors, including public and private entities, government bodies, and non-profits, it is intended for PII controllers or processors implementing data de-identification to enhance privacy protections (ISO, 2018a).

Importance of ISO/IEC 20889:

This document focuses on de-identification techniques commonly used for structured datasets and those organized in tabular form, such as those stored in key-value databases. These techniques may not be suitable for more complex datasets that include free-form text, images, audio, or video. While employing de-identification techniques is advisable to reduce the risk of re-identification, their effectiveness can vary. The document introduces the concept of a formal privacy measurement model to guide the application of these techniques (ISO, 2024b).

Principles of ISO/IEC 20889:

The principles-based framework offers a structured approach with various implementation options, including multiple deidentification techniques specified in ISO/IEC 20889. The

framework is organized into four key areas that address the environment and conditions under which deidentified data is accessed by users, as well as how the process and access are governed.

- Context Assessment: Assessing external information available to potential adversaries based on the circumstances of data availability. Administrative and technical controls are implemented to mitigate risks associated with context.
- Data Assessment: Evaluating how additional information accessible to adversaries could potentially reveal personal information. Restricting the data made available helps mitigate risks related to data exposure.
- Identifiability Assessment and Mitigation: Measuring the identifiability of data and assessing the likelihood and success of potential attacks based on contextual and data assessments. Mitigation strategies involve controls, data minimization, and transformation techniques aligned with established standards.
- Deidentification Governance: Implementing documented procedures and processes to ensure effective execution of assessments and mitigations. Establishing controls and response mechanisms manages risks throughout the lifecycle of deidentified data, from preparation to usage by authorized users.

Function of ISO/IEC 20889:

Specifically, ISO/IEC 20889 outlines terminology, categorizes de-identification techniques based on their attributes, and describes their effectiveness in mitigating the risk of re-identification (bsik.nowledge, 2018).

ix. ISO/IEC 27555

- This standard provides guidelines for deleting personal data and includes descriptions of different methods.
- This standard helps organizations develop policies and procedures for secure data deletion.

ISO/IEC 27555 overview:

This document provides comprehensive guidelines for developing and implementing policies and procedures for the deletion of personally identifiable information (PII) within organizations. It standardizes terminology for PII deletion, offers an efficient approach for defining deletion rules, and outlines the necessary documentation.

Moreover, it details the roles, responsibilities, and processes involved in PII deletion. Aimed at organizations that store or process PII, this document does not address specific legal provisions from national laws or contracts, specific deletion rules for particular PII categories defined by controllers, deletion mechanisms, the reliability, security, and suitability of these mechanisms, or specific techniques for data de-identification (bsi.knowledge, 2021).

Importance of ISO/IEC 27555:

This document details privacy-enhancing data de-identification techniques aligned with the privacy principles in ISO/IEC 29100. It specifies terminology, categorizes de-identification techniques based on their characteristics, and outlines their applicability for minimizing re-identification risk. Applicable to organizations of all types and sizes, including public and private companies, government entities, and non-profits, it is intended for PII controllers or processors acting on a controller's behalf to implement data de-identification processes to enhance privacy (ISO, 2018b).

Principle of ISO/IEC 27555:

ISO/IEC 27555 provides guidelines for the deletion of Personally Identifiable Information (PII) and aligns with the principles of ISO/IEC 29100, such as "data minimization" and "use, retention, and disclosure limitation." These principles emphasize reducing the amount of PII collected, limiting its use, and ensuring it is retained only for as long as necessary.

ISO/IEC 27555 supports these principles by offering a structured approach to PII deletion, ensuring that data is removed according to well-defined rules that comply with legal and regulatory requirements. This approach enhances data privacy and security by systematically eliminating unnecessary PII, thereby reducing the risk of data breaches and ensuring compliance with privacy laws and standards (ISO, 2021).

Function of ISO/IEC 27555:

This standard provides comprehensive guidance on the deletion of Personally Identifiable Information (PII) using a systematic approach that supports the "Privacy framework" outlined in ISO/IEC 29100. By following these guidelines, organizations can ensure that PII deletion processes are structured, efficient, and compliant with established privacy principles. This approach not only aligns with data minimization and retention limitations but also enhances overall data protection and privacy management (ISO, 2024c).

To conclude, these ISO standards play a crucial role in information security and privacy management within organizations. They provide frameworks, guidelines, and controls that help organizations implement privacy-preserving techniques to protect personal data from unauthorized access, use, or disclosure. By adhering to these standards, organizations can ensure compliance with regulatory requirements and enhance trust with stakeholders regarding their data protection practices.

8. Existing Frameworks for Privacy Preserving

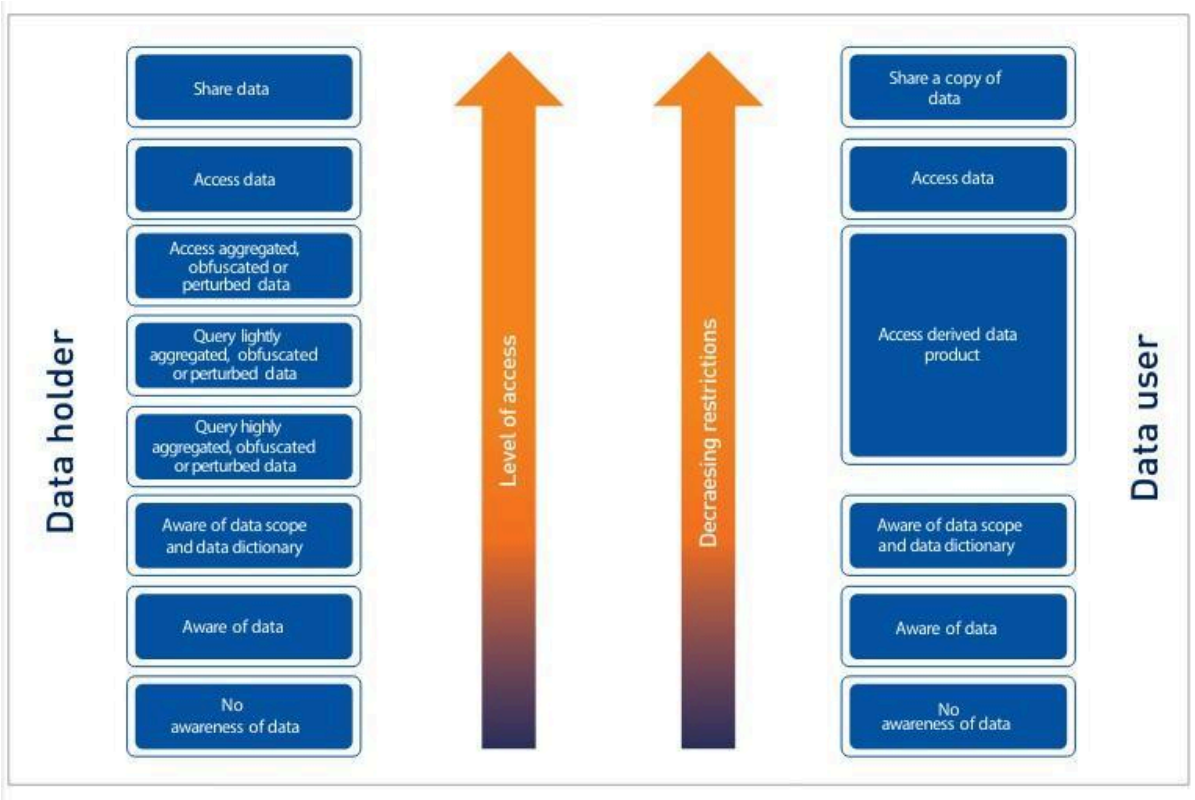
Australia: Governance dimensions for data sharing frameworks

For three years, the Australian Computer Society (ACS) has collaborated with many of Australia's top data scientists and companies to develop a practical, workable solution to the problem of enabling the benefits of smart services and data sharing while maintaining the fundamental right to privacy for citizens.

The major challenge for data scientists, government policymakers, and businesses that collect and use personal data is to enable the benefits of smart services and data sharing while safeguarding citizens' fundamental right to privacy. A general framework for data sharing and use is depicted in Figure 4, showing a range from the most restrictive to the least restrictive sharing methods.

It includes stages from not sharing any knowledge to sharing datasets. The framework incorporates various levels of access, such as awareness of data existence, access to derived data products, and querying data. Data products include metadata, provenance data, aggregated data, and modified data. This approach allows for different risk controls at various project phases (Opperman, 2019).

Figure 4. Framework for data sharing and use



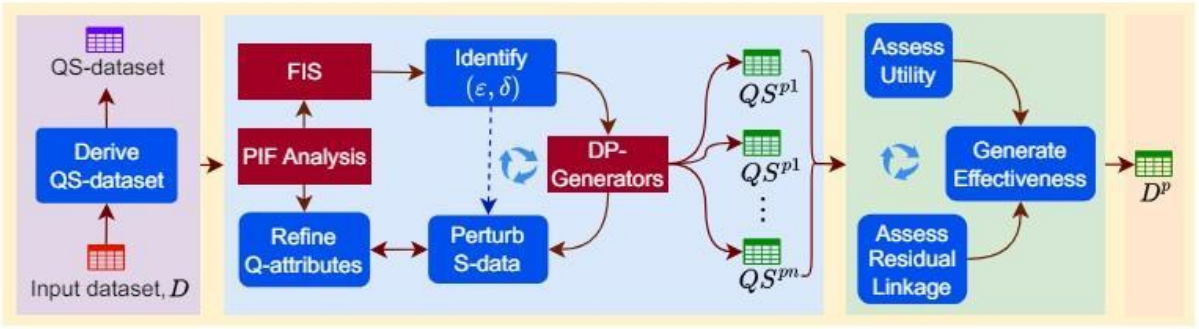
Source: (Opperman, 2019).

OptimShare: The Unified Frameworks for Privacy Preserving Data Sharing

Chamikara et al., (2023) introduce OptimShare, a utility-focused, multi-criteria solution designed to selectively perturb input datasets optimized for specific real-world applications. OptimShare integrates differential privacy, fuzzy logic, and probability theory to create a comprehensive tool for privacy-preserving data sharing. Empirical evaluations demonstrate that OptimShare effectively balances data utility and strong privacy, making it suitable for various real-world scenarios (Chamikara et al., 2023).

OptimShare is overseen by a central authority, such as a government agency, to enable Controlled Partially Perturbed Non-Interactive Data Sharing (CPNDS) with differential privacy, as illustrated in Figure 1, showing the modular arrangement of the OptimShare framework. D^p denotes the perturbed output dataset derived from the input dataset D. FIS stands for the Fuzzy Inference System, PIF represents the Personal Information Factor, and QS^{pi} indicates the intermediate perturbed instances of the QS dataset. The objective is to create a privacy-preserving version of the existing dataset for third-party analytics. To enhance security, user role management is implemented to regulate access levels.

Figure 8. Optimshare: a framework for privacy-preserving data sharing



Source: (Chamikara et al., 2023).

CREDENTIAL: A Framework for Privacy-Preserving Cloud-Based Data Sharing

Hörandner et al. (2016) proposed CREDENTIAL with the aim of developing and demonstrating innovative cloud-based services for securely storing, managing, and sharing identity information and personal data (Hörandner et al., 2016). By utilizing advanced cryptographic techniques such as proxy re-encryption and redactable signatures, CREDENTIAL seeks to enhance security in cloud identity management. Furthermore, it will improve password-based authentication by incorporating strong hardware-based multi-factor authentication mechanisms. The following sections outline CREDENTIAL's basic approach and principles (Blaze et al., 2006b; Johnson et al., 2002).

The CREDENTIAL consortium developed five key objectives to achieve its vision of privacy-preserving cloud-based identity management solutions:

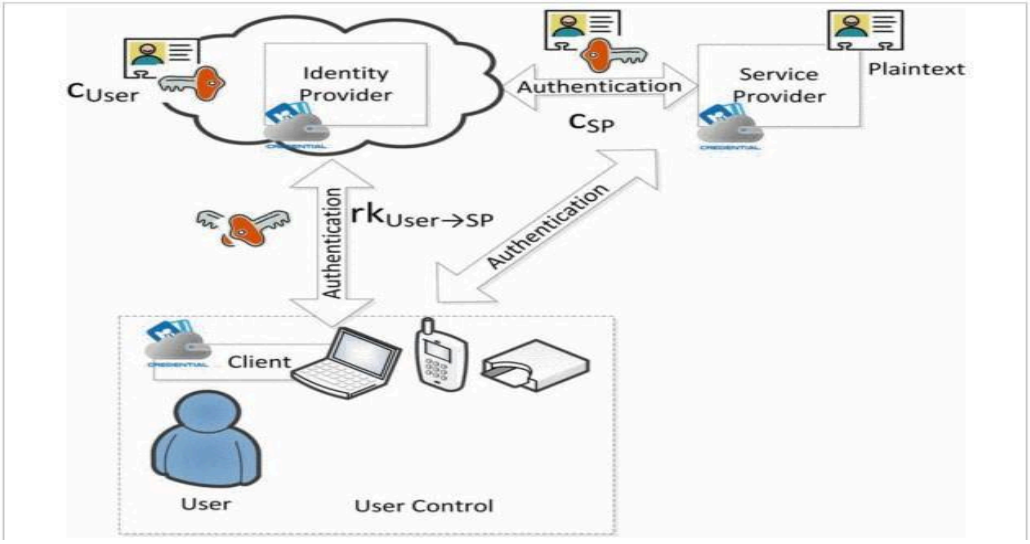
1. Adapt novel cryptographic mechanisms to securely store and share identity data in the cloud.
2. Enhance existing cloud authentication with strong hardware-based multi-factor mechanisms.
3. Design and develop a user-friendly, portable cloud identity, and data management architecture.
4. Create secure and portable implementations.
5. Evaluate and pilot the results in pre-production environments across various domains.

These cryptographic mechanisms will protect data confidentiality, integrity, and authenticity in the cloud while preserving user privacy through efficient re-encryption and selective disclosure features. Strong hardware-based multi-factor mechanisms will replace insecure password-based authentication. The developed platform and identity provider will adhere to modern security and privacy principles, using established standards and protocols

for easy integration with minimal modifications. Finally, CREDENTIAL's results will undergo extensive testing and evaluation in near-operational pilots across e-Government, e-Health, and e-Business domains (Hörandner et al., 2016).

Proxy re-encryption and redactable signatures address privacy challenges in cloud-based data sharing and identity management. Proxy re-encryption enables secure, end-to-end encrypted data sharing by allowing a semi-trusted proxy to transform ciphertexts without accessing the plaintext or decryption keys. Redactable signatures maintain data integrity and authenticity, enabling parts of a signed message to be removed without invalidating the signature (Blaze et al., 2006b; Johnson et al., 2002). Figure 9 depicts the fundamental architecture, highlighting all key stakeholders. Each stakeholder utilizes CREDENTIAL technology, making them CREDENTIAL-enabled.

Figure 9. CREDENTIAL basic architecture



Source: (Hörandner et al., 2016).

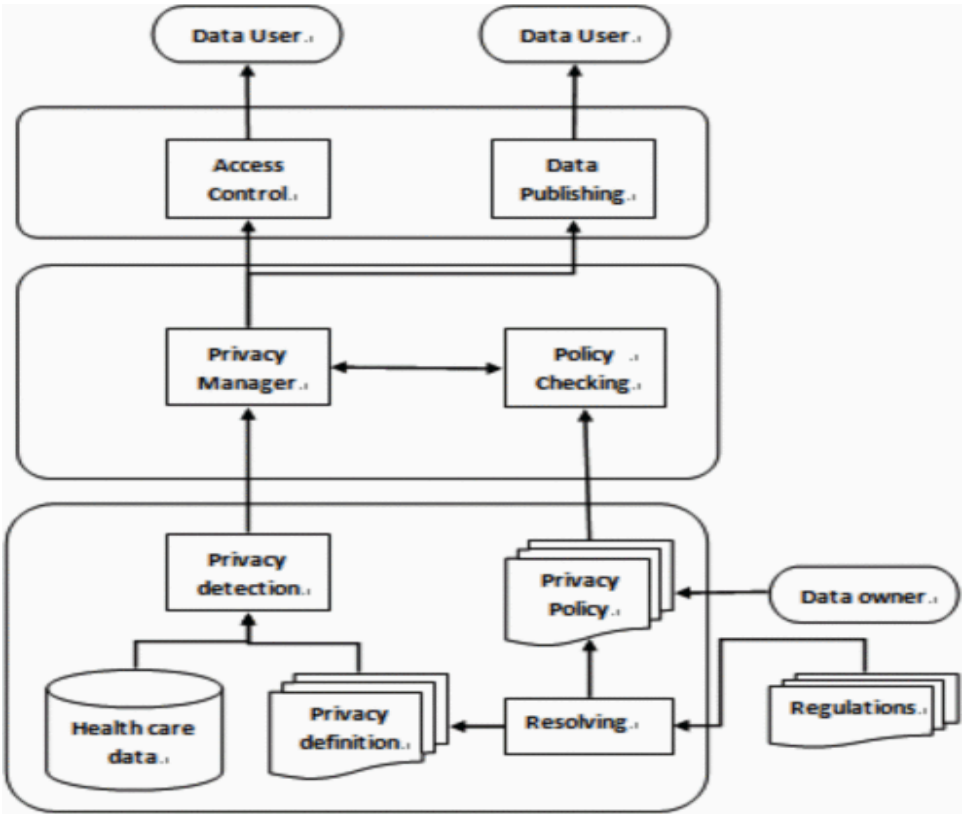
Framework for Privacy Preserving Health Care Data Sharing

Healthcare data is highly valuable for scientific research and analysis, leading to an increasing demand for sharing such data. However, since healthcare data often includes sensitive patient information, sharing it directly poses significant privacy risks. The privacy protection issues in healthcare data sharing can be divided into three categories: defining

privacy, managing privacy laws and regulations in information systems, and achieving privacy protection during data sharing based on defined privacy data and policies.

To address these issues, Chen et al., (2012) propose a three-layer framework for privacy-preserving healthcare data sharing, as illustrated in Figure 10. The bottom layer, forming the foundation, includes components for private data definition, privacy policy definition, and privacy detection, based on regulations such as Health Insurance Portability and Accountability Act (HIPAA) and R(97) (R(97) refers to the Council of Europe Recommendation No. R (97) 5 on the Protection of Medical Data. The middle layer enforces these policies, ensuring each record is suitable for sharing. The top layer provides direct services for data sharing through access control and publishing components (Lei Chen et al., 2012).

Figure 10. Framework for privacy-preserving healthcare data sharing



Source: (Lei Chen et al., 2012).

2.2 Ethical Considerations in Data Privacy

Ethical considerations play a crucial role in evaluating data privacy frameworks, particularly in ensuring vendor practices align with principles that safeguard personal data. Frameworks such as the GDPR and the DPDPA emphasize core values like transparency, consent, and fairness (GDPR.EU, 2024; Pop, 2023). However, these frameworks often fall short in addressing the ethical implications of emerging technologies, such as AI and big data, and the complex integration of vendor practices (Medium, 2023).

Floridi and Taddeo (2016) argue for aligning technological advancements with societal values (L. Floridi & Taddeo, 2016), while Brey, (2012) advocates for Value-Sensitive Design (VSD) in technology development (Brey, 2012). Despite their comprehensive, existing frameworks struggle to adapt to the dynamic nature of technology and ethical dilemmas it presents. There is a notable gap in developing guidelines that account for these evolving concerns and establishing robust ethical standards for new technological contexts.

2.3 Existing Ethical Frameworks

Several established ethical frameworks guide data sharing practices. The Belmont Report outlines principles such as respect for persons, beneficence, and justice, relevant to data sharing (Belmont Report, 1979). Similarly, the OECD Privacy Guidelines focus on limiting data collection, ensuring data accuracy, and providing access to data (OECD, 2023). While these frameworks provide a solid foundation, they face significant challenges in integrating emerging technologies, managing cross-border data sharing, and adopting to dynamic threats (L. Floridi & Taddeo, 2016; Vegesna, 2023).

The need for these frameworks to evolve is critical, as they must now consider the ethical implications of vendor practices and the complexities of global data flows. Enhancing these frameworks requires a dynamic and adaptive approach, integrating vendor practices, harmonizing cross-border standards, and engaging stakeholders to ensure ethically aligned data sharing.

2.4 Privacy-Preserving Data Sharing

Privacy-preserving data sharing techniques aim to balance privacy with effective data utilization. Techniques such as de-identification and anonymization are crucial in protecting sensitive information (Mishra, 2024). However, these techniques primarily focus on technical solutions, often overlooking the ethical considerations that should guid their application. For instance, while these methods protect privacy, they may also reduce data utility or inadvertently introduce biases.

Recent innovations, such as blockchain-based privacy solutions for AI (Hai & Liu, 2022), highlight the need for a multifaced approach that integrates ethical principles with technical advancements. There is a critical need to develop a framework that not only advances these technical solutions but also embeds ethical considerations, ensuring responsible and balanced data sharing practices.

2.5 Privacy-Preserving Techniques

Key privacy-preserving techniques, such as homomorphic encryption and federated learning, play a vital role in protecting data throughout its lifecycle. Homomorphic encryption allows computations on encrypted data without decryption, preserving privacy during data processing (Jiang et al., 2021). Differential privacy introduces noise to data queries, preventing the identification of individuals while still allowing useful data analysis (Dwork, 2006). Federated learning decentralized model training, reducing the risk of data breaches by keeping data local (Yazici et al., 2023).

However, while these techniques are advanced, they often fail to address the ethical implications of their implication, particularly concerning fairness and data utility. Current methods primarily focus on the technical aspects of privacy (Habbal et al., 2024; Vegesna, 2023), neglecting the broader ethical considerations that must be integrated to ensure balanced and fair data sharing practices. To bridge this gap, the proposed framework should advance these technical solutions while integrating ethical principles to ensure that privacy measures are balanced with fairness and utility in data sharing.

2.6 Data Anonymization

Data anonymization transforms personally identifiable information (PII) into non-identifiable formats, allowing data use for legitimate purposes while ensuring privacy (Secoda, 2024).

Techniques such as k-anonymity, generalization, masking, and t-closeness each have distinct advantages and limitations.

- **K-Anonymity**: Ensures each record is indistinguishable from at least k-1 others but may suffer from homogeneity and background knowledge attacks (Hussien et al., 2013; Samarati & Sweeney, 2007).
- **Generalization**: Replaces specific data values with broader categories, potentially leading to significant information loss (Mandal & Nigam, 2012; Mishra, 2024).
- **Masking**: Substitutes sensitive data with non-revealing formats, useful for adhering to privacy regulations (Imperva, 2024).
- **l-Diversity**: Enhances k-anonymity by ensuring diverse representation of sensitive attributes but can be labour-intensive and vulnerable to certain attacks (Jain et al., 2016; Li et al., 2007).

Advanced Techniques

- **T-Closeness**: Ensures the distribution of sensitive attributes within equivalence classes is similar to the overall dataset, reducing exposure risks (Li et al., 2007).
- **Bucketization**: Groups records into buckets to obscure links between sensitive and identifying information (K. Wang et al., 2016).
- **Slicing**: Partitions data horizontally and vertically, preserving attribute correlations and preventing membership disclosure (Bhuvanesh et al., 2017).

While these advanced techniques offer improved privacy protection, they do not fully address the ethical implications of data sharing, particularly in balancing privacy and data utility. The proposed framework should evaluate these techniques and propose solutions that integrate ethical considerations, ensuring that data management practices are both effective and ethically sound.

Table 2. provides a clear comparison between what current frameworks and techniques offer (strengths) and where they fall short (gaps) in addressing both ethical and technical challenges in privacy-preserving data sharing. The table illustrates that while there are strong ethical and technical frameworks in place for privacy-preserving data sharing, they often lack flexibility, adaptability, and integration with ethical considerations, particularly in the face of emerging technologies and complex data-sharing environments.

Table 2. Strengths and Gaps in Privacy-Preserving Frameworks and Techniques

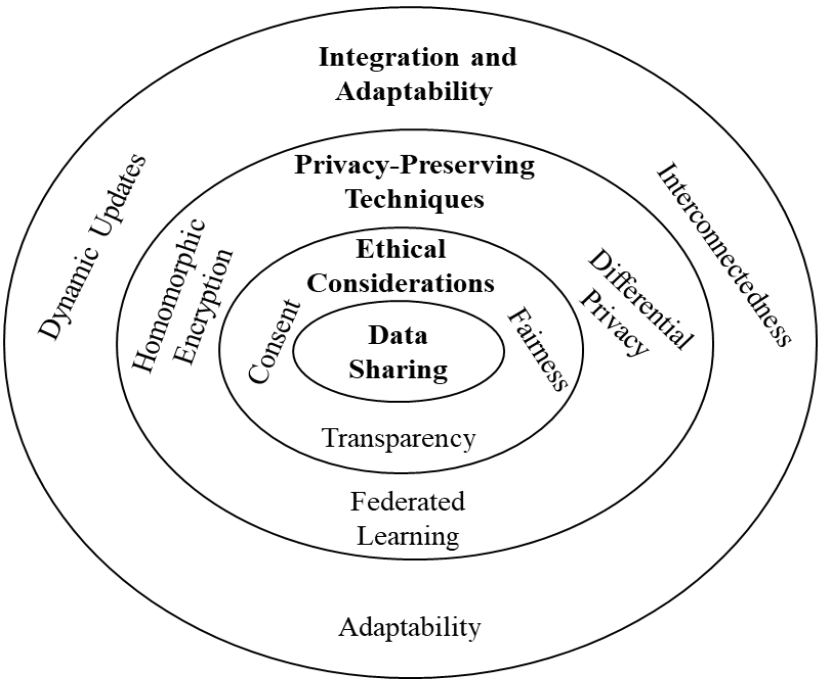
Aspect	Existing Frameworks/Techniques	Identified Gaps
Ethical Considerations	- GDPR and DPDPA emphasize transparency, consent, and fairness. - VSD integrates autonomy, justice, and beneficence in technology.	- Lack of adaptability to emerging technologies (e.g., AI, big data). - Insufficient focus on ethical implications of vendor practices.
Privacy-Preserving Data Sharing	- Techniques like de-identification, anonymization, and differential privacy protect sensitive information.	- Predominantly technical focus, with limited integration of ethical considerations. - Challenges in balancing privacy with data utility.
Privacy-Preserving Techniques	- Homomorphic encryption allows computation on encrypted data. - Federated learning decentralizes model training to enhance privacy.	- Overemphasis on technical solutions, overlooking fairness and ethical implications. - Difficulty in adapting to dynamic threats.
Data Anonymization Techniques	- K-anonymity and t-closeness ensure privacy by making data indistinguishable or balancing distribution of attributes.	- Susceptibility to specific attacks (e.g., homogeneity attacks) and significant information loss. - Lack of comprehensive ethical oversight.
Ethical Frameworks	- Belmont Report and OECD Guidelines emphasize principles like respect, beneficence, and accuracy.	- Struggles with integrating emerging technologies and managing cross-border data sharing. - Need for more dynamic and adaptable frameworks.

Figure 1 presents a structured approach to managing data privacy by integrating ethical principles with advanced technical solutions. The framework includes five key layers, (1) the Core Framework Layer, which establishes foundational principles; (2) Ethical Considerations, focusing on transparency, consent, fairness which evolving standards; (3) Privacy-Preserving Techniques, covering both basic and advanced methods, including homomorphic encryption, differential privacy, and federated learning; (4) Integration and Adaptability, ensuring responsiveness to new technologies and regulations, with dynamic updates and interconnectedness, (5) the outcome layer, aiming for robust data protection and responsible management. This framework provides a cohesive and dynamic model for balancing effective data protection with ethical practices.

The Comprehensive Data Sharing and Privacy-Preserving Framework is unique in its integrated approach, bridging the gap between ethical considerations and technical solutions in data management. Unlike traditional frameworks that often treat these aspects separately, this model combines them in a structured, layered format. It addresses the gaps in the literature by covering both basic and advanced privacy-preserving techniques, ensuring a comprehensive approach to data protection.

The framework's emphasis on adaptability allows it to remain relevant in the face of emerging technologies and evolving regulations, a feature commonly overlooked in existing models. By proactively addressing the ethical implications of new technologies, it ensures that data management practices are both effective and aligned with contemporary societal values, providing a dynamic and cohesive solution for modern data privacy challenges.

Figure 11. Holistic Data Sharing Framework



Source: By Authors.

9. Results

The analysis of current privacy-preserving frameworks and techniques reveals significant gaps in both ethical and technical aspects, particularly in the context of emerging technologies and complex data-sharing environments. The results are organized around three key areas, (1) the limitations of existing privacy-preserving technologies, (2) the inadequacies in ethical frameworks, and (3) the integration challenges of ethical principles with technical advancements.

First, the review of privacy-preserving technologies, such as k-anonymity, l-diversity, and differential privacy, highlights both their strengths in protecting data and their limitations in practical applications. K-anonymity ensures that each record is indistinguishable from at least k-1 others, making it a foundational technique in privacy protection (El Emam & Dankar, 2008; Sweeney, 2002). However, it is vulnerable to homogeneity and background knowledge attacks, which can compromise its effectiveness in protecting sensitive information (Q. Wang et al., 2011).

L-diversity improves upon k-anonymity by ensuring that sensitive attributes are well-represented within groups (Machanavajjhala et al., 2007), but it faces challenges related

to labour-intensive implementation and susceptibility to skewed distributions in the dataset (Mseer & Ahmed, 2024; Tamuhla et al., 2023).

Differential privacy introduces noise to data queries to prevent the identification of individuals, thus offering a robust privacy guarantee (Li et al., 2010). However, the added noise can significantly reduce data utility, particularly in small or sensitive datasets, making it difficult to strike a balance between privacy protection and data usability (X. Yang et al., 2017).

Technologies like t-closeness, bucketization, and slicing offer enhanced privacy protections by addressing specific vulnerabilities of earlier methods (Jayapradha & Prakash, 2022; Kumar et al., 2018). However, these approaches are primarily technical solutions and often fail to consider broader ethical implications, such as fairness, transparency, and user consent in data management (Sokolovska & Kocarev, 2018; Thapa & Camtepe, 2021). This limitation underscores the need for a more integrated approach that combines technical rigor with ethical considerations.

Second, the evaluation of ethical frameworks, such as the GDPR, DPDPA, and the Belmont Report, reveals that while they establish foundational principles for data privacy, they fall short in several critical areas. Ethical frameworks often struggle to keep pace with rapid technological advancements (Grover et al., 2024; Rigotti & Fosch-villaronga, 2024). For example, emerging technologies such as AI and blockchain introduce new ethical dilemmas, such as algorithmic bias and the opacity of AI decision-making processes (Owolabi et al., 2024), which existing frameworks are not fully equipped to handle.

This gap highlights the need for frameworks that are more adaptable to evolving technological landscapes. Current frameworks inadequately address the ethical implications of vendors adhere to ethical standards, particularly in global and complex data ecosystems .This gap suggests a need for more robust guidelines that ensure ethical consistency across the entire data lifecycle. The global nature of data sharing necessitates harmonized standards across jurisdictions. However, existing frameworks often lack the flexibility to address the complexities of cross-border data flows, leading to potential ethical and legal conflicts (Andanda & Mlotshwa, 2024). This challenge is particularly relevant in the context of international collaboration and data exchanges that require compliance with multiple, and sometimes conflicting, legal requirements.

Third, the results indicate that while there is a growing recognition of the need to integrate ethical principles into technical privacy-preserving solutions, significant challenges remain. Achieving a balance between privacy and data utility is a persistent challenge (N. Yuvaraj et al., 2022). Privacy-preserving techniques often prioritize privacy at the expense of data utility, which can limit the effectiveness of data-driven decision-making processes (Dhinakaran et al., 2024; Yuping Zhang et al., 2023).

There is a critical need for solutions that maintain privacy while preserving the usefulness of data, particularly in sectors like healthcare and finance where data accuracy and utility are paramount. Techniques such as homomorphic encryption and federated learning often promising privacy-preserving capabilities (L. Zhang et al., 2023). However, these technologies require further refinement to address ethical concerns such as fairness and transparency (Q. Yang, 2021). For instance, federated learning reduces the risk of data breaches by keeping data local, but it raises questions about fairness in model training and transparency in data usage (Bouacida & Mohapatra, 2021; Yifei Zhang et al., 2024). The results emphasize the importance of developing context-special ethical frameworks that consider the unique challenges of different data-sharing environments. These frameworks should integrate ethical reasoning with technical solutions to ensure that privacy-preserving measures are not only effective but also fair and transparent.

The analysis highlighted opportunities to enhance current frameworks by bridging the identified gaps. Specially, the integration of ethical principles into privacy-preserving technologies and the adoption of a more holistic approach to data management were identified as key areas for improvement. These results underscore the importance of developing a framework that not only preserves data privacy but also upholds ethical standards, thereby fostering responsible and transparent data sharing practices.

j) Comprehensive Ethical Data Sharing Framework

The Comprehensive Ethical Data Sharing Framework, as illustrated in Figure 2, addresses the critical need for a robust approach to data sharing that integrates ethical principles with privacy-preserving techniques. Central to this framework is the Holistic Ethical Framework, which ensures that data sharing practices are both technically effective and ethically sound. This central element emphasizes the balance between technical solutions and ethical

considerations, providing a foundation for all other components of the framework. It ensures that data management practices align with societal values and ethical standards.

The framework is organized along two main dimensions: Integration and Adaptability. Dynamic Adaptability is crucial for maintaining the framework's relevance in the face of rapidly evolving technologies and ethical challenges. This includes regular updates, continuous integration of new research, and feedback loops to refine and improve the framework. Integration of Vendor Practices ensures that third-party vendors comply with ethical standards through regular audits, ethical contracts, and training. Furthermore, Cross-Border Considerations address the complexities of international data sharing by developing harmonized standards and facilitating cooperation across jurisdictions to ensure consistent ethical practices.

Balancing Privacy and Utility is achieved through the application of privacy-preserving techniques such as homomorphic encryption, differential privacy, and federated learning, which protect privacy while allowing data to be useful. Addressing Evolving Threats and Contexts involves proactive measures to mitigate risks and adapt to new threats. Incorporating Ethical Principles, including transparency, accountability, fairness, beneficence, and nonmaleficence, ensures that the framework is ethically grounded. Stakeholder Engagement and Ensuring Accountability and Fairness further enhance the framework by incorporating diverse perspectives and maintaining high standards of responsibility.

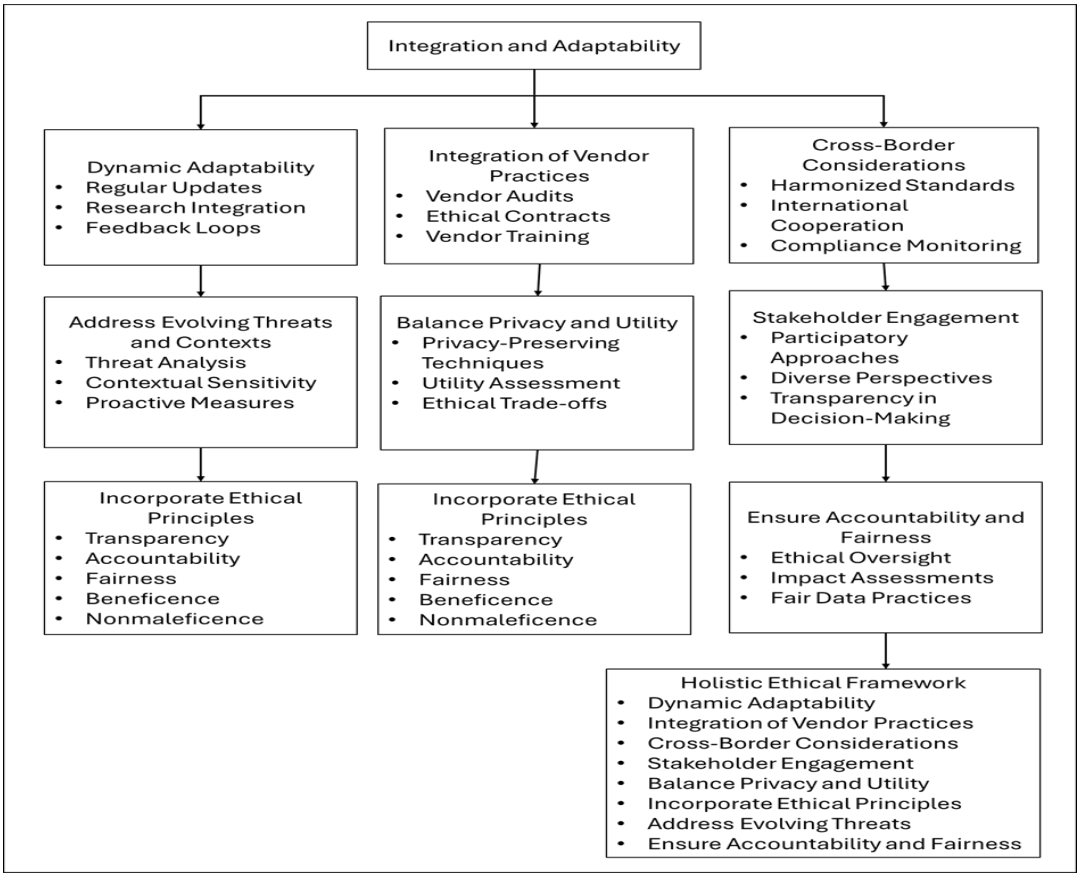
Together, these components create a comprehensive system that enhances data management and privacy protection effectively and ethically. Engaging with stakeholders ensures that the framework reflects diverse perspectives and addresses various concerns. This component involves participatory approaches that involve stakeholders in the decision-making process, considering input from various groups to ensure inclusivity, and ensuring transparency in decision-making. Engaging stakeholders helps to create a framework that is more responsive to the needs and values of different communities.

Ensuring Accountability and Fairness guarantees that data sharing practices are both fair and responsible. This involves establishing ethical oversight bodies or mechanisms to oversee compliance, conducting regular impact assessments to evaluate the effects of data sharing practices, and ensuring fair data practices that are just and equitable. This component ensures

that the framework maintains high standards of responsibility and fairness in data management.

This comprehensive framework integrates ethical principles and privacy-preserving techniques to create a robust system for data sharing. By addressing gaps in current methods and ensuring adaptability to new challenges, this framework aims to enhance data management and privacy protection in a manner that is both effective and ethically sound.

Figure 2. Comprehensive Ethical Data Sharing Framework



Source: By Authors.

10. Conclusion:

This literature survey provides a comprehensive examination of privacy-preserving data publishing (PPDP) and the associated techniques, principles, and ethical considerations necessary for maintaining data privacy in the digital age. By exploring various data anonymization techniques such as k-anonymity, generalization, and l-diversity, we identify

the strengths and limitations of each approach, emphasizing their applications and policy implications. Additionally, Differential Privacy offers robust privacy guarantees through methods like data perturbation, further enhancing data protection.

The survey outlines essential privacy-preserving principles, providing a foundation for understanding the ethical imperatives that guide privacy practices. The ethical foundations emphasize balancing principles in a technological era, which is critical to safeguarding individual privacy while promoting innovation. Recent advances in privacy-preserving techniques highlight ongoing developments and challenges in this dynamic field, while ISO/IEC standards underscore the importance of standardized approaches to privacy.

Finally, the proposed framework for privacy-preserving practices synthesizes the insights and methodologies discussed throughout the survey, providing a cohesive strategy for implementing effective privacy measures. This holistic approach balances the need for data utility with the imperative to protect individual privacy, ultimately contributing to the development of more secure and trustworthy data systems.

Given these considerations, it is imperative for each country to provide a comprehensive framework to address the pressing challenges of privacy in data sharing. This framework should integrate technical, ethical, and regulatory perspectives to ensure robust data privacy and security in the rapidly evolving digital landscape.

References:

6clicks. (2024a). *Glossary definition: Importance Of ISO/IEC 27005*. 6clicks. https://www.6clicks.com/resources/glossary/importance-of-iso-iec-27005

6clicks. (2024b). *ISO/IEC Standard: Definition & Benefits*. 6clicks. https://www.6clicks.com/resources/glossary/iso-iec-standard#:~:text=ISO%2FIEC standards are designed,they are produced or used.

6clicks. (2024c). *The expert's guide to ISO 27017*. 6clicks. https://www.6clicks.com/resources/guides/iso-27017#:~:text=ISO 27017 emphasizes the need,mechanisms based on data classification.

Ada Lovelace Institute. (2020). Examining the Black Box. In *DKUK*. https://www.adalovelaceinstitute.org/wp-content/uploads/2020/04/Ada-Lovelace-Institute-DataKind-UK-Examining-the-Black-Box-Report-2020.pdf

Agrawal, R., & Srikant, R. (2000). Privacy-preserving data mining. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, *29*(2), 439–450. https://doi.org/10.1145/335191.335438

Aldeen, Y. A. A. S., Salleh, M., & Razzaque, M. A. (2015). A comprehensive review on privacy preserving data mining. *SpringerPlus*, *4*(1), 1–36. https://doi.org/10.1186/s40064-015-1481-x

Algorithmic Justice League. (2024). *TECHNOLOGY SHOULD SERVE ALL OF US. NOT JUST THE PRIVILEGED FEW.* Algorithmic Justice League. https://www.ajl.org/

Ali, Z., & Bhaskar, S. B. (2016). Basic statistical tools in research and data analysis. *Indian Journal of Anaesthesia*, *60*(9), 662–669. https://doi.org/10.4103/0019-5049.190623

Allard, T., Nguyen, B., & Pucheral, P. (2014). METP: Revisiting privacy-preserving data publishing using secure devices. *Distributed and Parallel Databases*, *32*(2), 191–244. https://doi.org/10.1007/s10619-013-7122-x

Alliantist. (2024). *ISO/IEC 27017 Cloud Security Controls*. Isms.Online. https://www.isms.online/iso-27017/

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Article 29 Working Party (Art. 29 WP). (2014). Opinion 05/2014 on Anonymisation Techniques. *Papers of the Article 29 Working Party*, *0829/14/EN*(April), 1–37. http://ec.europa.eu/justice/data-protection/index_en.htm%0Ahttp://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

Aslam, B., & Zou, C. (2009). Distributed certificate and application architecture for VANETs. *Proceedings - IEEE Military Communications Conference MILCOM*, 1–7. https://doi.org/10.1109/MILCOM.2009.5379867

Ayala-Rivera, V., McDonagh, P., Cerqueus, T., & Murphy, L. (2014). A Systematic comparison and evaluation of k-Anonymization algorithms for practitioners. *Transactions on Data Privacy*, *7*(3), 337–370.

Bahrami, M., & Singhal, M. (2015). A light-weight permutation based method for data privacy in mobile cloud computing. *Proceedings - 2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, MobileCloud 2015*, 189–196. https://doi.org/10.1109/MobileCloud.2015.36

Banerjee, A., Hasan, M., Rahman, M. A., & Chapagain, R. (2017). CLOAK: A Stream Cipher Based Encryption Protocol for Mobile Cloud Computing. *IEEE Access*, *5*(March), 17678–17691. https://doi.org/10.1109/ACCESS.2017.2744670

Banerjee, S., & Sharma, S. (2019). Privacy concerns with Adhaar. *Communications of the ACM*, *62*(11), 80. https://doi.org/10.1145/3353770

Barocas, S. (2016). Big Data ' S Disparate Impact. *California Law Review*, *104*(671), 671–732.

Belsis, P., & Pantziou, G. (2012). A k-anonymity privacy-preserving approach in wireless medical monitoring environments. *Personal and Ubiquitous Computing*. https://doi.org/10.1007/s00779-012-0618-y

Bhatia, T., & Verma, A. K. (2017). Data security in mobile cloud computing paradigm: a survey,

taxonomy and open research issues. *Journal of Supercomputing*, *73*(6), 2558–2631. https://doi.org/10.1007/s11227-016-1945-y

Bhuvanesh, S., V N, A., Gladies, A. X. S., & M, V. (2017). Micro data Privacy Preserving Using Slicing. *Ijarcce*, *6*(3), 358–362. https://doi.org/10.17148/ijarcce.2017.6382

Binjubeir, M., Ahmed, A. A., Ismail, M. A. Bin, Sadiq, A. S., & Khurram Khan, M. (2019). Comprehensive survey on big data privacy protection. *IEEE Access*, *8*, 20067–20079. https://doi.org/10.1109/ACCESS.2019.2962368

Blaze, M., Bleumer, G., & Strauss, M. (2006a). Divertible protocols and atomic proxy cryptography. *Advances in Cryptology — EUROCRYPT'98*. https://link.springer.com/chapter/10.1007/BFb0054122

Blaze, M., Bleumer, G., & Strauss, M. (2006b). Divertible protocols and atomic proxy cryptography. In *Advances in Cryptology — EUROCRYPT'98* (pp. 127–144). Springer. https://link.springer.com/chapter/10.1007/BFb0054122

Boneh, D., & Franklin, M. (2001). Identity-based encryption from the weil pairing. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2139 LNCS*(3), 213–229. https://doi.org/10.1007/3-540-44647-8_13

Boudreaux, E., DeNardo, M. A., Denton, S. W., Sanchez, R., Feistel, K., & Dayalani, H. (2020). Privacy in P acy in Pandemic: Law andemic: Law, Technology echnology, and Public Health in the , and Public Health in the COVID-19 Crisis. In *RAND Corporation*. RAND Corporation. https://books.google.co.in/books?hl=en&lr=&id=TisTEAAAQBAJ&oi=fnd&pg=PP1&ots=FY4vJbyrrb&sig=SI_JrrNTdNkc9DnNR5o89OtkhKo&redir_esc=y#v=onepage&q&f=false

bsi.knowledge. (2021). *BS ISO/IEC 27555:2021 Information security, cybersecurity and privacy protection. Guidelines on personally identifiable information deletion*. Bsi.Knowledge. https://knowledge.bsigroup.com/products/information-security-cybersecurity-and-privacy-protection-guidelines-on-personally-identifiable-information-deletion?version=standard

bsik.nowledge. (2018). *BS ISO/IEC 20889:2018*. Bsi.Knowledgeknowledge. https://knowledge.bsigroup.com/products/privacy-enhancing-data-de-identification-terminology-and-classification-of-techniques?version=standard

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*, 77–91. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Bureau Veritas India. (2020). *KNOW WHY AN ORGANIZATION NEEDS TO BE ISO 27701 CERTIFIED*. Bureau Veritas India. https://www.bureauveritas.co.in/newsroom/know-why-organization-needs-be-iso-27701-certified#:~:text=ISO 27701 is a privacy,privacy regulations around the world.

Butcher, M. P. (2009). At the foundations of information justice. *Ethics and Information Technology*, *11*(1), 57–69. https://doi.org/10.1007/s10676-009-9181-2

Casas-Roma, J., Herrera-Joancomartí, J., & Torra, V. (2017). A survey of graph-modification techniques for privacy-preserving on networks. *Artificial Intelligence Review*, *47*(3), 341–366. https://doi.org/10.1007/s10462-016-9484-8

Chamikara, M. A. P., Jang, S. I., Oppermann, I., Liu, D., Ruj, S., Pal, A., Mohammady, M., Camtepe, S., Young, S., Dorrian, C., & David, N. (2023). OptimShare : A Unified Framework for Privacy Preserving Data Sharing – Towards the Practical Utility of Data with Privacy. *ArXiv.Org*, 1–15. https://doi.org/10.48550/arXiv.2306.03379

Chen, Lei, Yang, J. J., Wang, Q., & Niu, Y. (2012). A framework for privacy-preserving healthcare data sharing. *2012 IEEE 14th International Conference on E-Health Networking, Applications and Services, Healthcom 2012*, 341–346. https://doi.org/10.1109/HealthCom.2012.6379433

Chen, Liqun, & Urian, R. (2015). DAA-A: Direct Anonymous Attestation with Attributes. *International Conference on Trust and Trustworthy Computing*. https://doi.org/http://dx.doi.org/10.1007/978-3-319-22846-4_14

Chen, Y., & Cheung, A. S. Y. (2017). The Transparent Self Under Big Data Profiling: Privacy and Chinese Legislation on the Social Credit System. In *Special Issue Comparative Legal Reasoning: Essays in Honour of Geoffrey Samuel* (Vol. 12, Issue 2, pp. 356–378). The Journal of Comparative Law. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2992537

Clarin. (2024). *Principles of Data Processing*. Clarin. https://www.clarin.eu/content/principles-data-processing#:~:text=According to the principle of data minimisation%2C personal data shall,collected%2C stored or otherwise processed.

Cobb, M. (2024). *data masking*. TechTarget. https://www.techtarget.com/searchsecurity/definition/data-masking

Dabah, G. (2023). *What is Data Minimization? Main Principles & Techniques*. Piiano. https://www.piiano.com/blog/data-minimization

DataGuard. (2024). *ISO 27002: All you need to know about the standard*. DataGuard. https://www.dataguard.co.uk/blog/iso-27002/#three

de Fuentes, J. M., González-Manzano, L., Tapiador, J., & Peris-Lopez, P. (2017). PRACIS: Privacy-preserving and aggregatable cybersecurity information sharing. *Computers and Security*, *69*, 127–141. https://doi.org/10.1016/j.cose.2016.12.011

DIREC. (2023). *Accountability Privacy Preserving Computation via Blockchain*. Digital Research Centre Denmark. https://direc.dk/accountability-privacy-preserving-computation-via-blockchain/#:~:text=Privacy-preserving computation with accountability allows computation on private data,actors who tried to cheat.

DLA Piper. (2024). *DATA PROTECTION LAWS OF THE WORLD*. DLA Piper. https://www.dlapiperdataprotection.com/index.html?t=transfer&c=IN

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science*

Advances, *4*(1), 1–5. https://doi.org/10.1126/sciadv.aao5580

Drozd, O. (2017). Privacy Pattern Catalogue : A Tool for Integrating Privacy Principles of ISO / IEC 29100 into the Software Olha Drozd To cite this version : HAL Id : hal-01619742 Privacy Pattern Catalogue : A Tool for Integrating. In *HAL*. https://inria.hal.science/hal-01619742/document#:~:text=In the standard the requirements,and notice%3B individual participation and

Duncan, G. T., & Pearson, R. W. (1991). Enhancing access to microdata while protecting confidentiality: Prospects for the future. *Statistical Science*, *6*(3), 219–232. https://doi.org/10.1214/ss/1177011681

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography* (pp. 265–284). Springer. https://link.springer.com/chapter/10.1007/11681878_14

Dwork, C., & Roth, A. (2013). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, *9*(3–4), 211–487. https://doi.org/10.1561/0400000042

Emam, K. El. (2006). Data Anonymization Practices in Clinical Research. *Health (San Francisco)*, 1–15. https://www.ehealthinformation.ca/web/default/files/wp-files/2006-Data-Anonymization-Practices.pdf

European Medicines Agency. (2024). *Guidance on the details of the classification of variations requiring assessment according to Article 62 of Regulation (EU) 2019 / 6 for veterinary medicinal products and on the documentation to be submitted pursuant to those variations* (Vol. 31, Issue April 2023). https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/guidance-details-classification-variations-requiring-assessment-according-article-62-regulation-eu-2019-6-veterinary-medicinal-products-documentation-be-submitted-pursuant-those-variat

European Union. (2016). *Regulation 2016/679 of the European parliament and the Council of the European Union*. http://eur-lex.europa.eu/eli/reg/2016/679/oj

European Union. (2024). *International data transfers*. Edpb. https://www.edpb.europa.eu/sme-data-protection-guide/international-data-transfers_en

Executive Office of the President. (2015). Big Data and Differential Pricing. *Whitehouse.Gov*, *February*, 1–22. https://www.whitehouse.gov/sites/default/files/docs/Big_Data_Report_Nonembargo_v2.pdf

Exoscale. (2024). *ISO/IEC 27018:2019*. Exoscale. https://www.exoscale.com/compliance/iso27018/

F Pasquale. (2016). Black Box Society: The Secret Algorithms that Control Money and Information. In *Harvard University Press*. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061.c8

Faster Capital. (2024). *Privacy Preservation: The Importance of Data Anonymization Techniques*.

FasterCapital. https://fastercapital.com/content/Privacy-Preservation--The-Importance-of-Data-Anonymization-Techniques.html#:~:text=Suppression%3A Suppression%2C also known as,be anonymized through other means.

Fiesler, C., Boulder, Beard, N., & Maryland, B. C. (2020). No Robots, Spiders, or Scrapers: Legal and Ethical Regulation of Data Collection Methods in Social Media Terms of Service. *Proceedings of the International AAAI Conference on Web and Social Media*, *14*(1), 187–196. https://doi.org/https://doi.org/10.1609/icwsm.v14i1.7290

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Machine Learning and the City: Applications in Architecture and Urban Design*, *1*, 535–545. https://doi.org/https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Friedman, B., & Nissenbaum, H. (1996). User Autonomy: Who Should Control What and When? *Conference on Human Factors in Computing Systems - Proceedings*, 433. https://doi.org/10.1145/257089.257434

GDPR. (2016). *Recital 26 EU GDPR*. PrivazyPlan. https://www.privacy-regulation.eu/en/recital-26-GDPR.htm

GDPR. (2024a). *Art. 4 GDPRDefinitions*. Intersoft Consulting. https://gdpr-info.eu/art-4-gdpr/

GDPR. (2024b). *Recital 33 Consent to Certain Areas of Scientific Research**. Intersoft Consulting. https://gdpr-info.eu/recitals/no-33/

Ghafghazi, H., Mougy, A. El, Mouftah, H. T., & Adams, C. (2016). 11 - Security and Privacy in LTE-based Public Safety Network. In *identity based encryption* (pp. 317–364). Wireless Public Safety Networks 2. https://doi.org/https://doi.org/10.1016/B978-1-78548-052-2.50011-6

Ghinita, G., Kalnis, P., & Tao, Y. (2011). Anonymous Publication of Sensitive Transactional Data. *IEEE Transactions on Knowledge and Data Engineering*, *23*(2), 161–174. https://doi.org/https://doi.org/10.1109/TKDE.2010.101

Ghinita, G., Tao, Y., & Kalnis, P. (2008). On the anonymization of sparse high-dimensional data. In *Proceedings - International Conference on Data Engineering*. https://doi.org/10.1109/ICDE.2008.4497480

Ghorbel, A., Ghorbel, M., & Jmaiel, M. (2022). Accountable privacy preserving attribute-based access control for cloud services enforced using blockchain. *International Journal of Information Security*, *21*(3), 489–508. https://doi.org/10.1007/s10207-021-00565-4

Gope, P. (2019). LAAP: Lightweight anonymous authentication protocol for D2D-Aided fog computing paradigm. *Computers and Security*, *86*, 223–237.

https://doi.org/10.1016/j.cose.2019.06.003

Guan, Z., Zhang, Y., Wu, L., Wu, J., Li, J., Ma, Y., & Hu, J. (2019). APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT. *Journal of Network and Computer Applications*, *125*(August 2018), 82–92. https://doi.org/10.1016/j.jnca.2018.09.019

Hai, X., & Liu, J. (2022). PPDS: Privacy Preserving Data Sharing for AI applications Based on Smart Contracts. *Proceedings - 2022 IEEE 46th Annual Computers, Software, and Applications Conference, COMPSAC 2022*, 1561–1566. https://doi.org/10.1109/COMPSAC54236.2022.00248

Hamilton, R. H., & Sodeman, W. A. (2020). The questions we ask: Opportunities and challenges for using big data analytics to strategically manage human capital resources. *Business Horizons*, *63*(1), 85–95. https://doi.org/10.1016/j.bushor.2019.10.001

Hanna, K. T. (2024). *ISO 27001*. TechTarget. https://www.techtarget.com/whatis/definition/ISO-27001

Hao, K. (2021). *Deepfake porn is ruining women's lives. Now the law may finally ban it. After years of activists fighting to protect victims of image-based sexual violence, deepfakes are finally forcing lawmakers to pay attention.* MIT Technology Review. https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/

Harvey, A. (2024). *The Ultimate Guide to ISO 27002*. Isms.Online. https://www.isms.online/iso-27002/

Hay, M., Machanavajjhala, A., Miklau, G., Chen, Y., & Zhang, D. (2016). Principled evaluation of differentially private algorithms using DPBENCH. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, *26-June-20*, 139–154. https://doi.org/10.1145/2882903.2882931

He, D., Chan, S., & Guizani, M. (2016). An Accountable, Privacy-Preserving, and Efficient Authentication Framework for Wireless Access Networks. *IEEE Transactions on Vehicular Technology*, *65*(3), 1605–1614. https://doi.org/10.1109/TVT.2015.2406671

Healey, R. (2024). *Purpose Limitation Principle in Data Privacy Regulations: Your Guide*. Formiti. https://formiti.com/purpose-limitation-principle-in-data-privacy-regulations-your-guide/

HICKS, J. (2024). *Key Facts and Benefits of ISO 27018*. Schellman. https://www.schellman.com/blog/iso-certifications/key-facts-and-benefits-of-iso-27018

Hörandner, F., Krenn, S., Migliavacca, A., Thiemer, F., & Zwattendorfer, B. (2016). CREDENTIAL: A framework for privacy-preserving cloud-based data sharing. *Proceedings - 2016 11th International Conference on Availability, Reliability and Security, ARES 2016*, 742–749. https://doi.org/10.1109/ARES.2016.79

Hussien, A. A., Hamza, N., & Hefny, H. A. (2013). Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing. *Journal of Information Security*, *4*(2). https://doi.org/10.4236/jis.2013.42012

i-SCOOP. (2024). *Personal data breach notification and communication duties under the GDPR*. I-SCOOP. https://www.i-scoop.eu/gdpr/personal-data-breach-notification/

IEEE. (2017). Ethically Aligned Design: Version 2 - For Public Discussion. *IEEE Standards*, 1–263. https://standards.ieee.org/industry-connections/ec/ead-v1/

Imine, Y., Lounis, A., & Bouabdallah, A. (2020). An accountable privacy-preserving scheme for public information sharing systems. *Computers and Security*, *93*. https://doi.org/10.1016/j.cose.2020.101786

Imperva. (2024). *Anonymization*. Imperva. https://www.imperva.com/learn/data-security/anonymization/#:~:text=from the data.-,Data Anonymization Techniques,*" or "x".

Information Commissioner's Office. (2022). Chapter 5 : technologies (PETs). In *Informa* (Issue September). ico. https://ico.org.uk/media/aboutthe-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf

Information with Insight. (2024). *Data Anonymisation and L-Diversity*. Information with Insight. https://informationwithinsight.com/2019/03/12/data-anonymisation-and-l-diversity/

ISO. (2018a). *ISO/IEC 20889:2018 Privacy enhancing data de-identification terminology and classification of techniques*. ISO. https://www.iso.org/standard/69373.html

ISO. (2018b). ISO/IEC 20889 Privacy enhancing data deidentification terminology and classification of techniques. In *Globalspec*. https://standards.globalspec.com/std/10267481/iso-iec-20889

ISO. (2019). *ISO/IEC 27018:2019(en) Information technology — Security techniques — Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors*. ISO. https://www.iso.org/obp/ui/en/#iso:std:iso-iec:27018:ed-2:v1:en

ISO. (2021). *ISO/IEC 27555:2021(en) Information security, cybersecurity and privacy protection — Guidelines on personally identifiable information deletion*. ISO. https://www.iso.org/obp/ui/#iso:std:iso-iec:27555:ed-1:v1:en

ISO. (2023). *What are the Differences Between Anonymisation and Pseudonymisation*. Privacy Company. https://www.privacycompany.eu/blog/what-are-the-differences-between-anonymisation-and-pseudonymisation

ISO. (2024a). *Information technology — Security techniques — Information security risk management*. ISO. https://www.iso.org/obp/ui/#iso:std:iso-iec:27005:ed-2:v1:en

ISO. (2024b). *ISO/IEC 20889:2018(en) Privacy enhancing data de-identification terminology and classification of techniques*. ISO. https://www.iso.org/obp/ui/#iso:std:iso-iec:20889:ed-1:v1:en

ISO. (2024c). *ISO/IEC 27555:2021 — Information security, cybersecurity and privacy protection — Guidelines on personally identifiable information deletion (first edition)*. ISO. https://www.iso27001security.com/html/27555.html

ISO. (2024d). *ISO/IEC 29100:2011 Information technology — Security techniques — Privacy*

framework. ISO. https://www.iso.org/standard/45123.html#:~:text=Abstract,describes privacy safeguarding considerations%3B and

IT Governance UK. (2024). *ISO 27005*. IT Governance UK. https://www.itgovernance.co.uk/iso27005#:~:text=ISO 27005 is the international,organisation's ISO 27001 compliance project.

ITA. (2024). *5–DATA INTEGRITY AND PURPOSE LIMITATION*. Data Privacy Framework Program. https://www.dataprivacyframework.gov/framework-article/5–DATA-INTEGRITY-AND-PURPOSE-LIMITATION

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, *3*(1). https://doi.org/10.1186/s40537-016-0059-y

Jayapradha, J., Prakash, M., Alotaibi, Y., Khalaf, O. I., & Alghamdi, S. A. (2022). Heap Bucketization Anonymity - An Efficient Privacy-Preserving Data Publishing Model for Multiple Sensitive Attributes. *IEEE Access*, *10*, 28773–28791. https://doi.org/10.1109/ACCESS.2022.3158312

Jiang, L., Chen, L., Giannetsos, T., Luo, B., Liang, K., & Han, J. (2019). Toward Practical Privacy-Preserving Processing Over Encrypted Data in IoT: An Assistive Healthcare Use Case. *IEEE Internet of Things Journal*, *6*(6), 10177–10190. https://doi.org/10.1109/JIOT.2019.2936532

Johnson, R., Molnar, D., Song, D., & Wagner, D. (2002). Homomorphic Signature Schemes. In *Topics in Cryptology — CT-RSA 2002* (pp. 244–262). Springer. https://link.springer.com/chapter/10.1007/3-540-45760-7_17

Jonas, S., Siewert, S., & Cord Spreckelsen. (2019). Privacy-Preserving Record Grouping and Consent Management Based on a Public-Private Key Signature Scheme: Theoretical Analysis and Feasibility Study. *Journal of Medical Internet Research*, *21*(4). https://doi.org/10.2196/12300

Kak, A. (2020). *Regulating Biometrics: Global Approaches and Urgent Questions*. AI Now Institute. https://ainowinstitute.org/regulatingbiometrics.pdf

Kamara, S., Papamanthou, C., & Roeder, T. (2012). Dynamic searchable symmetric encryption. *Proceedings of the ACM Conference on Computer and Communications Security*, 965–976. https://doi.org/10.1145/2382196.2382298

Kaminska, J. (2024). *Cross-Border Data Transfer: How to Move Data Across Jurisdictions Without Headaches*. Anonos. https://www.anonos.com/blog/cross-border-data-transfer

Karyda, M., & Mitrou, L. (2016). Data Breach Notification: Issues and Challenges for Security Management. *Association for Information Systems*. http://aisel.aisnet.org/mcis2016%0Ahttp://aisel.aisnet.org/mcis2016/60

Khalil, I., Khreishah, A., & Azeem, M. (2014). Consolidated Identity Management System for secure mobile cloud computing. *Computer Networks*, *65*, 99–110. https://doi.org/10.1016/j.comnet.2014.03.015

Kirvan, P. (2024). *ISO 27002 (International Organization for Standardization 27002)*. TechTarget.

https://www.techtarget.com/searchsecurity/definition/ISO-27002-International-Organization-for-Standardization-27002#:~:text=ISO 27002 provides guidance on,identify potential threats and vulnerabilities.

Klein, L., van Deursen, N., Buchanan, W. J., Duff, A., Ottens, M., Kroes, P., Franssen, M., Van De Poel, I., de Bruijn, H., Herder, P. M., Yu, E., & Hutchison, D. (2002). Microdata Protection through Noise Addition. In *Inference Control in Statistical Databases* (Vol. 2316, pp. 97–116). Springer, Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/3-540-47804-3_8

Kokolakis, G., & Fouskakis, D. (2009). Importance partitioning in micro-aggregation. *Computational Statistics and Data Analysis*, *53*(7), 2439–2445. https://doi.org/10.1016/j.csda.2008.09.028

Kostka, G., Steinacker, L., & Meckel, M. (2021). Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science*, *30*(6), 671–690. https://doi.org/10.1177/09636625211001555

Kosutic, D. (2024). *What is ISO 27001? A quick and easy explanation*. Advisera. https://advisera.com/27001academy/what-is-iso-27001/

Kumar Singh, A., P. Keer, N., & Motwani, A. (2014). A Review of Privacy Preservation Technique. *International Journal of Computer Applications*, *90*(3), 17–20. https://doi.org/10.5120/15554-4239

Levich, S. K., Rokach, L., Elovici, Y., Member, I., & Shapira, B. (2010). Efficient Multidimensional Suppression for K-Anonymity. *IEEE Transactions on Knowledge and Data Engineering*, *22*, 334–347. https://doi.org/10.1109/TKDE.2009.91

Li, B., & He, K. (2023). Local generalization and bucketization technique for personalized privacy preservation. *Journal of King Saud University - Computer and Information Sciences*, *35*(1), 393–404. https://doi.org/10.1016/j.jksuci.2022.12.008

Li, F., Zou, X., Liu, P., & Chen, J. Y. (2011). New threats to health data privacy. *BMC Bioinformatics*, *12 Suppl 1*(Suppl 12). https://doi.org/10.1186/1471-2105-12-S12-S7

Li, N., Li, T., & Venkatasubramanian, S. (2007). t-Closeness : Privacy Beyond k-Anonymity and-Diversity t -Closeness : Privacy Beyond k -Anonymity and -Diversity. *IEEE*, *July*, 106–115. https://doi.org/10.1109/ICDE.2007.367856

Li, T. C. (2021). Privacy in P acy in Pandemic: Law andemic: Law, Technology echnology, and Public Health in the , and Public Health in the COVID-19 Crisis. *Loyola University Chicago Law Journal*, *52*(3). https://lawecommons.luc.edu/luclj/vol52/iss3/5?utm_source=lawecommons.luc.edu%2Fluclj%2Fvol52%2Fiss3%2F5&utm_medium=PDF&utm_campaign=PDFCoverPages

Li, T., Li, N., Zhang, J., & Molloy, I. (2012). Slicing: A new approach for privacy preserving data publishing. *IEEE Transactions on Knowledge and Data Engineering*, *24*(3), 561–574. https://doi.org/10.1109/TKDE.2010.236

Liang, K., Au, M. H., Liu, J. K., Susilo, W., Wong, D. S., Yang, G., Yu, Y., & Yang, A. (2015). A secure and efficient Ciphertext-Policy Attribute-Based Proxy Re-Encryption for cloud data sharing. *Future Generation Computer Systems*, *52*, 95–108. https://doi.org/10.1016/j.future.2014.11.016

Lin, C., Shen, Z., Chen, Q., & Sheldon, F. T. (2016). A Data Integrity Verification Scheme in Mobile Cloud Computing. *Journal of Network and Computer Applications*. https://doi.org/http://dx.doi.org/10.1016/j.jnca.2016.08.017

Liu, J. K., Au, M. H., Susilo, W., & Zhou, J. (2012). Enhancing Location Privacy for Electric Vehicles (at the Right time). In *Computer Security – ESORICS 2012* (pp. 397–414). Springer. https://link.springer.com/chapter/10.1007/978-3-642-33167-1_23

Loukil, F., Ghedira-Guegan, C., Boukadi, K., & Benharkat, A. N. (2018). LIoPY: A Legal Compliant Ontology to Preserve Privacy for the Internet of Things. *Proceedings - International Computer Software and Applications Conference*, *2*, 701–706. https://doi.org/10.1109/COMPSAC.2018.10322

M Sravanthi, & Bhaskar, N. (2016). Shared Authority Based Privacy-Preserving Authentication Protocol in Cloud Computing. *Internation Journal of Sceince Engineering and Technology Research*, *5*(22). https://www.researchgate.net/publication/349138842_Shared_Authority_Based_Privacy-Preserving_Authentication_Protocol_in_Cloud_Computing

Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). ℓ-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data*, *1*(1). https://doi.org/10.1145/1217299.1217302

Majeed, A., & Lee, S. (2016). Anonymization Techniques for Privacy Preserving Data Publishing: A Comprehensive Survey. *IEEE Access*, *9*, 8512–8545. https://doi.org/10.1109/ACCESS.2020.3045700

Mandal, A., & Nigam, M. K. (2012). A Review On Data Anonymization Technique For Data Publishing. *International Journal of Engineering Research & Technology (IJERT)*, *1*(10), 46–54. https://d1wqtxts1xzle7.cloudfront.net/64342587/a-review-on-data-anonymization-technique-for-data-publishing-IJERTV1IS10210-libre.pdf?1599121871=&response-content-disposition=inline%3B+filename%3DIJERT_A_Review_On_Data_Anonymization_Tec.pdf&Expires=1716968

Mishra, A. (2024). Privacy-Preserving Data Sharing Platform. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *8*(4). https://doi.org/10.55041/IJSREM32225

Mishra, S. (2023). *Key Principles of 27701 Certification*. SlideShare. https://www.slideshare.net/slideshow/key-principles-of-27701-certification/260819533

Mivule, K. (2014). An Investigation Of Data Privacy And Utility Using Machine Learning As A Gauge. *3619387*, *October*, 17–33. http://pqdtopen.proquest.com/pubnum/3619387.html

Mollah, M. B. (2018). Security and Privacy Challenges in Cloud Computing. *Proceedings of the 2018 Cyber Resilience Conference, CRC 2018*, *84*(September 2016), 38–54. https://doi.org/10.1109/CR.2018.8626872

Mollah, M. B., Azad, M. A. K., & Vasilakos, A. (2017). Security and privacy challenges in mobile cloud computing: Survey and way ahead. *Journal of Network and Computer Applications*, *84*(September 2016), 38–54. https://doi.org/10.1016/j.jnca.2017.02.001

Mozur, P. (2019). *One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority*. The New York Times. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

Nabeel, M., Shang, N., & Bertino, E. (2013). Privacy preserving policy-based content sharing in public clouds. *IEEE Transactions on Knowledge and Data Engineering*, *25*(11), 2602–2614. https://doi.org/10.1109/TKDE.2012.180

National Academies of Sciences, Engineering, and M. (2017). Practice 7: Respect for the Privacy and Autonomy of Data Providers. In C. CF (Ed.), *Principles and Practices for a Federal Statistical Agency: Sixth Edition*. National Academies Press (US). National Academies of Sciences, Engineering, and Medicine; Division of Behavioral and Social Sciences and Education; Committee on National Statistics

National Research Council. (1984). Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines. In T. R. Jabine TB, Straf ML, Tanur JM (Ed.), *Report of the Advanced Research Seminar on Cognitive Aspects of Survey Methodology* (Committee). National Academy Press.

National Research Council. (2008). *Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics Programs Using the American Community Survey for the National Science Foundation's Science and Engineering Workforce Statistics Programs*. https://nap.nationalacademies.org/catalog/12244/using-the-american-community-survey-for-the-national-science-foundations-science-and-engineering-workforce-statistics-programs

National Research Council. (2023). Protecting Participants and Facilitating Social and Behavioral Sciences Research. *Committee on National Statistics and Board on Behavioral, Cognitive, and Sensory Sciences*. https://doi.org/https://doi.org/10.17226/10638

National Science and Technology Council. (2023). NATIONAL STRATEGY TO ADVANCE PRIVACY-PRESERVING DATA SHARING AND ANALYTICS. *FAST-TRACK ACTION COMMITTEE ON ADVANCING PRIVACY-PRESERVING DATA SHARING AND ANALYTICS NETWORKING AND INFORMATION TECHNOLOGY RESEARCH AND DEVELOPMENT SUBCOMMITTEE of the NATIONAL SCIENCE AND TECHNOLOGY COUNCIL*, *March*. https://www.nitrd.gov/about/

Nicanfar, H., Hosseininezhad, S., Talebifard, P., & Leung, V. C. M. (2013). Robust privacy-preserving authentication scheme for communication between Electric Vehicle as Power Energy Storage and power stations. *2013 IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2013*, 55–60. https://doi.org/10.1109/INFCOMW.2013.6562908

Nietzsche, F. (2021). *Thus Spake Zarathustra A Book for All and None*. Thomas Common. https://www.gutenberg.org/files/1998/1998-h/1998-h.htm

Nissenbaum, H. . (2018). *Stop thinking about consent: It isn't possible and it isn't right*. Harvard Bus Review. https://store.hbr.org/product/stop-thinking-about-consent-it-isn-t-possible-and-it-isn-t-right/H04J77

Nolasco, F. (2023). *WHAT EXACTLY IS THE ISO 29100 PRIVACY FRAMEWORK?* Datatunnel. https://datatunnel.io/2023/01/06/what-exactly-is-the-iso-29100-privacy-framework/#:~:text=ISO%2FIEC 29100%3A2011 establishes,Describes privacy safeguarding considerations.

NQA. (2021). *ISO 27701*. NQA GLOBAL CERTIFICATION BODY. https://www.nqa.com/en-me/certification/standards/iso-27701#:~:text=ISO 27701%2C also abbreviated as,Processors to manage data privacy.

oananiculaescu. (2017). *k-Anonymity and cluster based methods for privacy*. Elf11.Github.Io. https://elf11.github.io/2017/05/22/kanonymity.html

OIPC. (2014). *OIPC – Privacy by Design Resources*. OIPC. https://iapp.org/resources/article/oipc-privacy-by-design-resources/

Olakanmi, O. O., & Dada, A. (2019). An Efficient Privacy-preserving Approach for Secure Verifiable Outsourced Computing on Untrusted Platforms. *International Journal of Cloud Applications and Computing*, *9*(2), 79–98. https://doi.org/https://doi.org/10.4018/IJCAC.2019040105

OneTrust. (2022). *Privacy 101: Data Transfers*. OneTrust DataGuidance. https://www.dataguidance.com/resource/privacy-101-data-transfers

Opperman, I. (2019). *Privacy-Preserving Data Sharing Frameworks* (Issue December). https://www.infogovanz.com/wp-content/uploads/2020/01/191202-ACS-Privacy-eReport.pdf

Palo Alto Networks. (2024). *What Is Data Privacy?* Palo Alto Networks. https://www.paloaltonetworks.com/cyberpedia/data-privacy#:~:text=Purpose limitation in data privacy refers to the principle that,%2C explicit%2C and legitimate purposes.

Park, I. S., Lee, Y. D., & Jeong, J. (2013). Improved identity management protocol for secure mobile cloud computing. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 4958–4965. https://doi.org/10.1109/HICSS.2013.262

Pasquale, F. (2020). *New Laws of Robotics*. Harvard University Press. https://doi.org/https://doi.org/10.4159/9780674250062

Pika, A., Wynn, M. T., Budiono, S., Hofstede, A. H. M. T., van der Aalst, W. M. P., & Reijers, H. A. (2020). Privacy-preserving process mining in healthcare. *International Journal of Environmental*

Research and Public Health, *17*(5). https://doi.org/10.3390/ijerph17051612

Podesta, J., Pritzker, P., Moniz, E. J., Holdren, J., & Zients, J. (2015). Big Data: Seizing Opportunities, Preserving Values - Interim Progress Report. In *Executive Office of the President of USA* (Issue February). https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

Polat, H., & Du, W. (2003). Electrical Engineering and Computer Science Privacy-Preserving Collaborative Filtering Using Randomized Perturbation Techniques Privacy-Preserving Collaborative Filtering using. In *SURFACE*. https://surface.syr.edu/eecs/18?utm_source=surface.syr.edu%2Feecs%2F18&utm_medium=PDF&utm_campaign=PDFCoverPages

Polyakova, A., & Meserole, C. (2019). Exporting digital authoritarianism: The Russian and Chinese models. *Policy Brief, Democracy and Disorder Series*, 1–22. https://www.brookings.edu/wp-content/uploads/2019/08/FP_20190827_digital_authoritarianism_polyakova_meserole.pdf

Pulls, T. (2012). Privacy-Preserving Transparency-Enhancing Tools. In *Karlstad University Faculty of Health, Science and Technology Department of Mathematics and Computer Science*.

Qlik. (2024). *Data Mining*. Qlik. https://www.qlik.com/us/data-analytics/data-mining#:~:text=The key types of data,ensemble methods%2C and text mining.

Qu, Y., Nosouhi, M. R., Cui, L., & Yu, S. (2019). Chapter 6 - Privacy Preservation in Smart Cities. In *Smart Cities Cybersecurity and Privacy* (pp. 75–88). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-815032-0.00006-8

Rachel, B. (2017). *Big Data Meets Big Brother as China Moves to Rate Its Citizens*. WIRED. https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion

Rahman, M., Paul, M. K., & Sattar, A. H. M. S. (2023). Efficient perturbation techniques for preserving privacy of multivariate sensitive data. *Array*, *20*(May), 100324. https://doi.org/10.1016/j.array.2023.100324

Ram Mohan Rao, P., Murali Krishna, S., & Siva Kumar, A. P. (2018). Privacy preservation techniques in big data analytics: a survey. *Journal of Big Data*, *5*(1). https://doi.org/10.1186/s40537-018-0141-8

Rana, M. E., & Jayabalan, M. (2016). Privacy Preserving Anonymization Techniques for Patient Data: An Overview. *3rd International Conference on Knowledge, Information and S Oftware Engineering (ICKIS2016)*, *January 2016*. https://www.researchgate.net/publication/318720516

Ranganathan, N. (2019). The Economy (and Regulatory Practice) That Biometrics Inspires: A Study of the Aadhaar Project. In *Regulating Biometrics: Global Approaches and Urgent Questions*. https://ainowinstitute.org/wp-content/uploads/2023/09/regulatingbiometrics-ranganathan.pdf

Ravichandran, K., Gavrilovska, A., & Pande, S. (2015). PiMiCo: Privacy Preservation via Migration in Collaborative Mobile Clouds. *Hawaii International*. https://doi.org/10.1109/HICSS.2015.628

Re, R. M., Solow-Niederman, A., Bussel, D., Brennan-Marquez, K., Kaminski, M., Levy, K., Minow, M., Parson, T., Schwartz, J., Tasioulas, J., & Volokh, E. (2019). Developing Artificially Intelligent Justice. *Stanford Technology Law Review 242*, *22*, 243–288. https://perma.cc/9DQ6-MH7E.

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). *Algorithmic impact assessments: A practical framework for public agency accountability.* In AI Now Institute. https://ainowinstitute.org/aiareport2018.pdf

Richman, A. (2023). *Advantages and Disadvantages of Anonymized Data*. K2view. https://www.k2view.com/blog/anonymized-data/#:~:text=Masking data is one of,Information)%2C and intellectual property.

Rivest, R. L., Adleman, L., & Dertouzos, M. L. (1978). ON DATA BANKS AND PRIVACY HOMOMORPHISMS. *Journal of Pidgin and Creole Languages*, *15*(2), 313–337. https://doi.org/10.1075/jpcl.15.2.04lef

Roberts, S. R. (2021). *The war on the Uyghurs*. manchesterhive. https://doi.org/https://doi.org/10.7765/9781526166258

Robinson, N., Graux, H., Botterman, M., & Valeri, L. (2009). Review of the European Data Protection Directive. In *Rand Europe Technical Report* (Issue January). https://ico.org.uk/media/about-the-ico/documents/1042349/review-of-eu-dp-directive.pdf

Rottondi, C., Fontana, S., & Verticale, G. (2014). A Privacy-Friendly Framework for Vehicle-to-Grid Interactions. In *Smart Grid Security* (pp. 1–15). Springer. http://dx.doi.org/10.1007/978-3-319-10329-7_1

Sahai, A., & Waters, B. (2005). Fuzzy identity-based encryption. *Lecture Notes in Computer Science*, *3494*, 457–473. https://doi.org/10.1007/11426639_27

Salas, J., & Domingo-Ferrer, J. (2018). Some Basics on Privacy Techniques, Anonymization and their Big Data Challenges. *Mathematics in Computer Science*, *12*(3), 263–274. https://doi.org/10.1007/s11786-018-0344-6

Salem, F. M., Ibrahim, M. H., & I.I., I. (2010). Non-interactive authentication scheme providing privacy among drivers in vehicle-to-vehicle networks. *6th International Conference on Networking and Services, ICNS 2010, Includes LMPCNA 2010; INTENSIVE 2010*, 156–161. https://doi.org/10.1109/ICNS.2010.28

Samarati, P. (2011). k-Anonymity. In *Encyclopedia of Cryptography and Security*. Springer. https://link.springer.com/referenceworkentry/10.1007/978-1-4419-5906-5_754#:~:text=Suppression can be applied at the tuple level (TS)%2C,the other cells in the

Samarati, P., & Sweeney, L. (2007). Protecting Privacy When Disclosing Information: K Anonymity and its Enforcement Through Suppression. *IEEE Xplore*.

Scytale. (2024). *ISO 27018*. Scytale. https://scytale.ai/glossary/iso-27018/#:~:text=The key principles and requirements of ISO 27018 are as,%2C data protection%2C and compliance.

Secoda. (2024). *What is Data Anonymization?* Secoda. https://www.secoda.co/glossary/what-is-data-anonymization#:~:text=Data anonymization serves the purpose,individuals associated with the data.

Shamir, A. (1984). Identity-Based Cryptosystems and Signature Schemes. *Advances in Cryptology*. https://doi.org/https://doi.org/10.1007/3-540-39568-7_5

Shao, J., CaoXiaohui, Z., Lin, LinHuang, & Xiaohui, L. (2010). Proxy re-encryption with keyword search. *Information Sciences*, *180*(13). https://doi.org/http://dx.doi.org/10.1016/j.ins.2010.03.026

Shen, J., Zhou, T., Chen, X., Li, J., & Susilo, W. (2018). Anonymous and Traceable Group Data Sharing in Cloud Computing. *IEEE Transactions on Information Forensics and Security*, *13*(4), 912–925. https://doi.org/10.1109/TIFS.2017.2774439

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems*, *10*(4). https://doi.org/10.1145/3419764

Singh, A. P. (2023). A Review of Privacy Preserving Data Publ. *International Journal of Emerging Research in Management &Technology*, *January*.

Sinjilawi, Y. K., AL-Nabhan, M. Q., & Abu-Shanab, E. A. (2014). Addressing security and privacy issues in cloud computing. *Journal of Emerging Technologies in Web Intelligence*, *6*(2), 192–199. https://doi.org/10.4304/jetwi.6.2.192-199

SISA. (2024). *Comparison between ISO 27005, OCTAVE & NIST SP 800-30*. SISA. https://www.sisainfosec.com/blogs/comparison-between-iso-27005-octave-nist-sp-800-30-sisa-blog/

Slijepčević, D., Henzl, M., Daniel Klausner, L., Dam, T., Kieseberg, P., & Zeppelzauer, M. (2021). k-Anonymity in practice: How generalisation and suppression affect machine learning classifiers. *Computers and Security*, *111*, 102488. https://doi.org/10.1016/j.cose.2021.102488

Sloot, B. van der. (2013). From Data Minimization to Data Minimummization. In *Discrimination and Privacy in the Information Society* (pp. 273–287). Springer. https://link.springer.com/chapter/10.1007/978-3-642-30487-3_15

Song, D. X., Wagner, D., & Perrig, A. (2000). Practical techniques for searches on encrypted data. *Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy*, 44–55. https://doi.org/10.1109/secpri.2000.848445

Sriharsha, A. V., & Parthasarathy, C. (2015). Perturbing sensitive data using additive noise. *International Journal of Applied Engineering Research*, *10*(17), 38296–38301.

Stevens, G. (2012). Data security breach notification laws. *Data Security Breaches: Notification Laws, Legislation and Identity Theft*, 1–29.

Sucasas, V., Mantas, G., Saghezchi, F. B., Radwan, A., & Rodriguez, J. (2016). An autonomous privacy-preserving authentication scheme for intelligent transportation systems. *Computers and Security*, *60*, 193–205. https://doi.org/10.1016/j.cose.2016.04.006

Sultan, S., & Jensen, C. D. (2021). Ensuring purpose limitation in large‐scale infrastructures with provenance‐enabled access control. *Sensors*, *21*(9). https://doi.org/10.3390/s21093041

Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowldege-Based Systems*, *10*(5), 571–588. https://doi.org/10.1142/S021848850200165X

Torkzadehmahani, R., Nasirigerdeh, R., Blumenthal, D. B., Kacprowski, T., List, M., Matschinske, J., Spaeth, J., Wenke, N. K., & Baumbach, J. (2022). Privacy-Preserving Artificial Intelligence Techniques in Biomedicine. *Methods of Information in Medicine*, *61 (Suppl*, e12–e27. https://doi.org/10.1055/s-0041-1740630

Tosoni, L. (2020). Article 4(5). Pseudonymisation Get access Arrow. In *The EU General Data Protection Regulation (GDPR): A Commentary* (pp. 132–137). Axford Academic. https://doi.org/https://doi.org/10.1093/oso/9780198826491.003.0011

Touceda, D. S., Cámara, J. M. S., Zeadally, S., & Soriano, M. (2015). Attribute-based authorization for structured peer-to-peer (P2P) networks. *Computer Standards & Interfaces*, *11*(1), 1–5. https://doi.org/http://dx.doi.org/10.1016/j.csi.2015.04.007

Tsai, J. Y., Egelman, S., Cranor, L., & Acquisti, A. (2010). The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research*, *22*(2). https://doi.org/https://doi.org/10.1287/isre.1090.0260

U.S. Department of Health and Human Services. (2017). *Office for Human Research Protections*. https://www.hhs.gov/ohrp/index.html

U.S. Department of Health, E. and W. (1973). *The Code of Fair Information Practices. Secretary's advisory committee on automated personal data systems, records, computers, and the rights of citizens viii*. DTIC. https://apps.dtic.mil/sti/citations/tr/ADA002177

U.S. Office of Management and Budget. (2014). *Guidance for providing and using administrative data for statistical purposes*. OMB Memorandum M-14-06.

Unal, D., Bennbaia, S., & Catak, F. O. (2022). Machine learning for the security of healthcare systems based on Internet of Things and edge computing. In *Cybersecurity and Cognitive Science* (Issue January, pp. 299–320). https://doi.org/10.1016/B978-0-323-90570-1.00007-3

United Nations. (2024). *Universal Declaration of Human Rights*. UN. https://www.un.org/en/about-us/universal-declaration-of-human-rights

Vallent, T. F., Hanyurwimfura, D., & Mikeka, C. (2021). Efficient certificate-less aggregate signature scheme with conditional privacy-preservation for vehicular ad hoc networks enhanced smart grid system. *Sensors*, *21*(9). https://doi.org/10.3390/s21092900

Vinogradov, S., & Pastsyak, A. (2012). Evaluation of Data Anonymization Tools. *DBKDA 2012 : The Fourth International Conference on Advances in Databases, Knowledge, and Data Applications*, *c*, 163–168. http://www.thinkmind.org/index.php?view=article&articleid=dbkda_2012_7_20_30085

Von Grafenstein, M. (2018). *The Principle of Purpose Limitation in Data Protection Laws: The Risk-based Approach, Principles, and Private Standards as Elements for Regulating Innovation*.

Wadhwa, P. (2024). *An In-Depth Guide to ISO 27017*. Sprinto. https://sprinto.com/blog/iso-27017/

Wang, Changji, Li, W., Li, Y., & Xu, X. (2013). A Ciphertext-Policy Attribute-Based Encryption Scheme Supporting Keyword Search Function. In *Cyberspace Safety and Security*. Springer.

Wang, Cong, Cao, N., Ren, K., & Lou, W. (2012). Enabling secure and efficient ranked keyword search over outsourced cloud data. *IEEE Transactions on Parallel and Distributed Systems*, *23*(8), 1467–1479. https://doi.org/10.1109/TPDS.2011.282

Wang, Cong, Wang, Q., Ren, K., & Lou, W. (2010). Privacy-preserving public auditing for data storage security in cloud computing. *Proceedings - IEEE INFOCOM*, 1–9. https://doi.org/10.1109/INFCOM.2010.5462173

Wang, H., Qin, B., Wu, Q., X., & L., Domingo-Ferrer, J. (2015). TPP: Traceable Privacy-Preserving Communication and Precise Reward for Vehicle-to-Grid Networks in Smart Grids. *IEEE Transactions on Information Forensics and Security*, *10*(11), 2340–2351. https://doi.org/10.1109/TIFS.2015.2455513

Wang, K., Wang, P., Fu, A. W., & Wong, R. C. W. (2016). Generalized bucketization scheme for flexible privacy settings. *Information Sciences*, *348*, 377–393. https://doi.org/10.1016/j.ins.2016.01.100

Wang, Q., Wang, C., Ren, K., Lou, W., & Li, J. (2011). Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing. *IEEE Transactions on Parallel and Distributed Systems*, *22*(5), 847–859. https://doi.org/http://dx.doi.org/10.1109/TPDS.2010.183

Wang, S., Zhou, J., Liu, J. K., Yu, J., Chen, J., & Xie, W. (2016). An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing. *IEEE Transactions on Information Forensics and Security*, *11*(6), 1265–1277. https://doi.org/10.1109/TIFS.2016.2523941

Westin, A. F. (1968). Privacy And Freedom. In *Washington and Lee Law Review* (Vol. 166, Issue 1). https://doi.org/: https://scholarlycommons.law.wlu.edu/wlulr/vol25/iss1/20

Wilson, R. L., & Rosen, P. A. (2023). Protecting Data through Perturbation Techniques: The Impact on Knowledge Discovery in Databases. *Journal of Database Management*, *14*(2), 4–26. https://doi.org/10.4018/978-1-59140-471-2.ch003

Yang, K., & Jia, X. (2013). An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing. *IEEE Transactions on Parallel and Distributed Systems*, *24*(9), 1717–1726. https://doi.org/http://dx.doi.org/10.1109/TPDS.2012.278

Yu, S., Wang, C., Ren, K., & Lou, W. (2010). Achieving secure, scalable, and fine-grained data access control in cloud computing. *Proceedings - IEEE INFOCOM*, 1–9. https://doi.org/10.1109/INFCOM.2010.5462174

Zhang, A., Bacchus, A., & Lin, X. (2016). Consent-based access control for secure and privacy-preserving health information exchange. *Security and Communication Networks*, *9*(16),

3496–3508. https://doi.org/10.1002/sec.1556

Zhang, J., Chen, B., Zhao, Y., Cheng, X., & Hu, F. (2018). Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues. *IEEE Access*, *6*(Idc), 18209–18237. https://doi.org/10.1109/ACCESS.2018.2820162

Zhao, C., Zhao, S., Zhao, M., Chen, Z., Gao, C. Z., Li, H., & Tan, Y. an. (2019). Secure Multi-Party Computation: Theory, practice and applications. *Information Sciences*, *476*, 357–372. https://doi.org/10.1016/j.ins.2018.10.024

Zhou, L., Varadharajan, V., & Hitchens, M. (2013). Achieving Secure Role-Based Access Control on Encrypted Data in Cloud Storage. *IEEE Transactions on Information Forensics and Security*, *12*(8). https://doi.org/http://dx.doi.org/10.1109/TIFS.2013.2286456