Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi
International Journal of Engineering Research and Development

UMAGD, (2023) 15(2), 574-583.

10.29137/umagd.1272085

Cilt/Volume:15 Sayı/Issue:2 Haziran/June 2023

Research Article

An Enterprise Data Privacy Governance Model: Security-Centric Multi-Model Data Anonymization

Yağmur Şahin[1] *, İbrahim Doğru[1]

[1] *Gazi University, Graduate School of Natural and Applied Sciences, Department of Information Security Engineering, 06560, Ankara, TÜRKİYE*
[2] *Gazi University, IoTLab, Department of Computer Engineering, 06500, Ankara, TÜRKİYE*

Başvuru/Received: *29/03/2023* ***Kabul / Accepted***: *15/04/2023* ***Çevrimiçi Basım / Published Online:*** *30/06/2023*

Son Versiyon/Final Version: *30/06/2023*

Abstract

The increasing need for data privacy and the rising complexity of data environments necessitate robust data anonymization techniques to safeguard personal and sensitive information. A multi-model approach to data anonymization can strike an optimal balance between privacy protection and data utility, integrating techniques such as data masking, differential privacy, machine learning algorithms, blockchain technology, and data encryption. This article introduces a Security-Centric Enterprise Data Anonymization Governance Model, a structured framework for managing data privacy across healthcare, finance, and government industries. The model ensures adherence to best practices and compliance with legal and regulatory requirements. The article addresses challenges in implementing data anonymization techniques, including maintaining data utility and preventing re-identification, by advocating for a multi-model approach that combines various technologies and methods. We suggest that by adopting this holistic approach, organizations can enhance their data protection measures and foster a culture of data privacy.

Key Words

“Data privacy, data anonymization, multi-model approach, data utility, data governance”

***Corresponding Author:** ygmrrshnn@gmail.com*

1.Introduction

Institutions in a variety of industries, including healthcare, finance, and government, deal with sensitive data concerning individuals and entities. The data is frequently used for research and analysis, and the insights gained from such analysis can be useful in shaping policy decisions and driving innovation.

Nonetheless, the use of these data raises significant privacy concerns, and institutions must take the necessary precautions to protect sensitive information from unauthorized access or misuse. Among these measures is the utilization of data anonymization techniques.

Recent studies indicate that data breaches are on the rise, with an increase of 62% in 2020 compared to 2022, resulting in 1,862 data compromises and approximately 422.1 million individuals affected in the United States alone (Statista, 2023). This highlights the urgency for organizations to adopt robust data privacy measures to safeguard sensitive information.

Various methods, including surveys, interviews, and data analytics, are used by organizations dealing with sensitive data. These techniques provide organizations with valuable insights into individuals and entities, enabling them to better comprehend trends and behaviors. However, these methods also present significant risks in the data collection phase and in regard to protection.

Data privacy through anonymization is a critical aspect of information management, and institutions must take proactive steps to ensure data is adequately protected. While data analysis is vital for institutional growth, protecting individual privacy is equally important.

Data anonymization is an approach to safeguarding data privacy, which involves removing identifying information from data sets while maintaining usability.

In this article, we will explore a security-centric enterprise data privacy governance model and explain how this approach can protect sensitive information while still allowing to facilitate data utility and data analysis.

The security-centric data anonymization governance model is a framework that outlines the various steps involved in managing data privacy. It encompasses everything from defining data privacy requirements to implementing technical and administrative controls. This multi-model provides a structured approach to data privacy management, ensuring that institutions adhere to best practices and meet legal and regulatory requirements.

We suggest that by implementing this model, enterprises can establish a culture of data privacy and integrate data protection into all aspects of their operations.

This paper is organized as follows: Section 2 offers a background overview; Section 3 examines relevant literature; Section 4 presents the proposed methodology; Section 5 reports the results; Section 6 showcases a hypothetical scenario employing a multi-model approach; Section 7 engages in a discussion of the study's outcomes; Section 8 addresses the study's limitations and provides recommendations for future research; and finally, Section 9 concludes the paper.

2.Background

Anonymization can be achieved through various methods, including aggregation, generalization, and masking (Lubarsky, 2010; Samarati & Sweeney, 1998).

While data anonymization is an important tool for protecting privacy, it presents significant obstacles that must be overcome to ensure its efficacy. The difficulty of completely removing all identifying information from a dataset is one of the greatest obstacles to data anonymization. In some cases, it may be possible to identify individuals based on seemingly innocuous information, such as their age, gender, or geographic location (El Emam et al., 2011). To address this challenge, data anonymization techniques must be carefully designed to ensure that all identifying information is removed or altered in a way that prevents re-identification.

Removing or altering data can make it less useful for analysis and may result in incomplete or inaccurate results. Therefore, data anonymization techniques must be carefully designed to balance privacy protection with data quality and utility when anonymizing data. It is important that anonymized data sets are still usable for analysis while protecting the confidentiality of the data.

To achieve this, organizations can use various techniques, such as differential privacy, which allows for more precise control over the amount of information that is removed or altered (Nyugen, 2019).

Additionally, the studies show anonymized datasets are still vulnerable to data breaches, which can compromise the privacy of individuals even if their identifying information has been removed (Ohm, 2009). Therefore, data anonymization techniques must be combined with robust data security measures, such as encryption and access controls, to ensure that anonymized data is protected from unauthorized access.

To overcome these challenges and ensure effective data anonymization, organizations must adopt a comprehensive approach to data privacy management (Kalloniatis et al., 2013).

For these and similar reasons, clearly defined boundaries, approaches, and multi-model solutions are needed.

3.Related Works

Multi-model applications refer to software applications that can work with multiple data models and sources simultaneously (Lu & Holubová, 2019).

In the context of the Enterprise Data Privacy Governance Model and data anonymization, multi-model applications can provide a powerful tool for managing complex data environments while protecting the privacy of individuals.

One of the key benefits of multi-model applications in the context of data privacy is their ability to aggregate and anonymize data from multiple sources. This is particularly important for organizations that collect data from a variety of different sources, as it allows them to consolidate this data and apply anonymization techniques uniformly across all sources (Jiang & Torra, 2022).

Multi-model applications for enterprise data privacy governance models and data anonymization could include a combination of techniques and technologies that work together to protect and anonymize sensitive data appropriately. If we need to list the general lines of these approaches:

- Data masking and tokenization: These techniques involve replacing sensitive data with fictitious values or tokens to anonymize it. A multi-model approach might combine these techniques with other data anonymization methods, such as data perturbation or aggregation (Tachepun & Thammaboosadee, 2020).

- Differential privacy: This is a privacy-enhancing technique that involves adding random noise to data so that it becomes more difficult to identify individuals. A multi-model approach might use differential privacy in combination with other techniques to achieve greater privacy protection (Nyugen, 2019).

- Machine learning algorithms: These can be used to automatically identify and classify sensitive data in a dataset. A multi-model approach might use a combination of supervised and unsupervised machine learning algorithms to identify and protect sensitive data (Mahmood & Jusas, 2022).

- Blockchain technology: This can be used to create a tamper-proof and transparent record of data access and usage, which can help ensure data privacy and compliance. A multi-model approach might combine blockchain technology with other techniques to create a comprehensive data governance model (Kshetri, 2018).

- Data encryption: This is a method of encoding data so that it can only be accessed by authorized parties. A multi-model approach might use encryption in combination with other techniques to provide a multi-layered approach to data privacy (Muhasin et al., 2016).

We suggest that combining these approaches with anonymization techniques further strengthens their effectiveness. It is essential to recognize that each approach possesses its own unique advantages and limitations, and the optimal balance achieved through their integration creates a better effective defense mechanism.

4.Methodology

Data anonymization is the process of removing or obscuring identifying information from datasets, thereby protecting individuals' privacy and reducing the risk of data breaches.

The data anonymization lifecycle involves several stages, including data collection, preparation, transformation, and evaluation. Data anonymization can be applied using many methods, such as k-anonymity, l-diversity, and t-closeness, each of which provides different levels of protection and utility for the data (Sweeney, 2002; Machanavajjhala et al., 2007).

In light of related research, we divided the data anonymization lifecycle into several key stages:

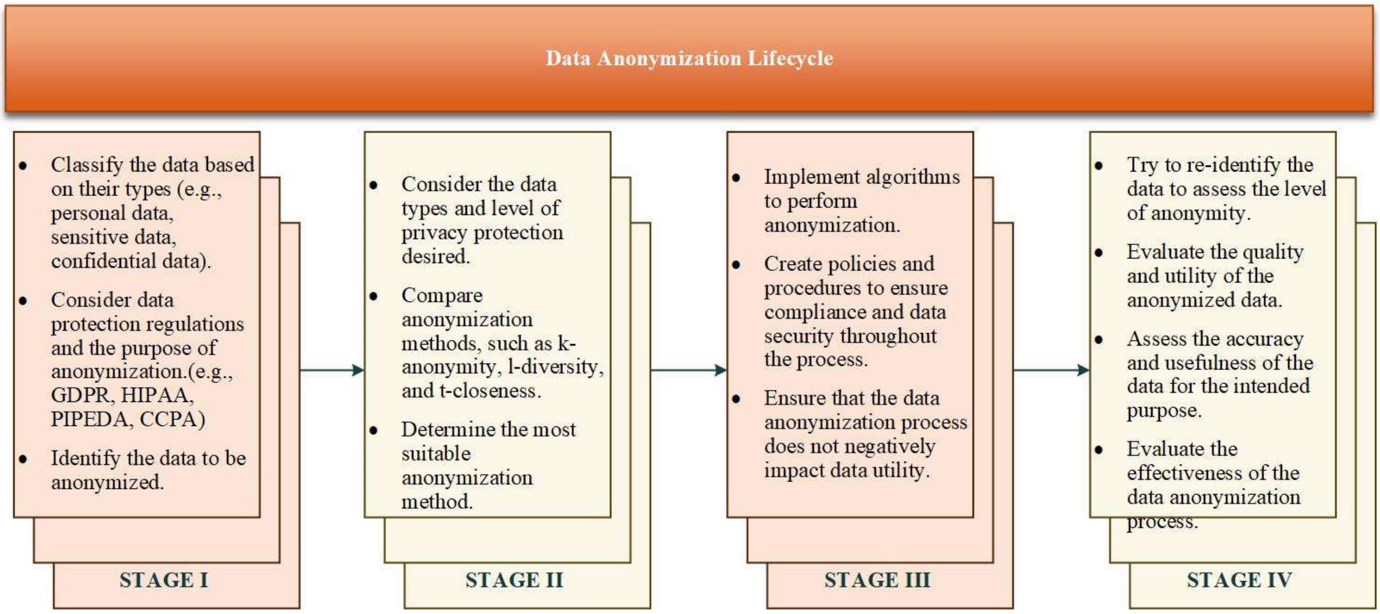
Stage I: The first stage entails identifying the data that must be anonymized, including the types of data and specific fields containing identifying information. This stage necessitates a thorough understanding of the data and its application and familiarity with any applicable data protection regulations.

Stage II: The second stage of the data anonymization lifecycle entails determining the best anonymization method. There are several methods for anonymizing data, including k-anonymity, l-diversity, and t-closeness, each of which provides varying levels of data protection and utility (Zhou & Pei, 2011). Enterprises must select the appropriate method based on the specific data to be anonymized and the level of privacy protection desired (LeFevre et al., 2006).

Stage III: The implementation of the anonymization process is the third stage of the data anonymization lifecycle. This may entail creating and implementing algorithms to perform anonymization, as well as creating policies and procedures to ensure data security throughout the process (Cavoukian & Jonas, 2011). In this stage, enterprises must also ensure that the anonymization process does not have a negative impact on the data's utility, as this can reduce the data's usefulness for analysis and decision-making (Mackey et al., 2016).

Stage IV: The final stage is data evaluation, in which organizations attempt to re-identify individuals from anonymized data to test the effectiveness of the anonymization process (Mourby et al., 2018). This stage entails evaluating the anonymized data's quality and utility, such as the accuracy of statistical analysis and the data's usefulness for the intended purpose (Xu et al., 2006).

The data anonymization lifecyle illustration is presented below to show how these stages appear in practice:



- Classify the data based on their types (e.g., personal data, sensitive data, confidential data).
- Consider data protection regulations and the purpose of anonymization.(e.g., GDPR, HIPAA, PIPEDA, CCPA)
- Identify the data to be anonymized.

- Consider the data types and level of privacy protection desired.
- Compare anonymization methods, such as k-anonymity, l-diversity, and t-closeness.
- Determine the most suitable anonymization method.

- Implement algorithms to perform anonymization.
- Create policies and procedures to ensure compliance and data security throughout the process.
- Ensure that the data anonymization process does not negatively impact data utility.

- Try to re-identify the data to assess the level of anonymity.
- Evaluate the quality and utility of the anonymized data.
- Assess the accuracy and usefulness of the data for the intended purpose.
- Evaluate the effectiveness of the data anonymization process.

Figure 1. Data anonymization lifecycle stages

When anonymizing data, organizations must also comply with data protection regulations such as GDPR (General Data Protection Regulation) and CCPA (California Privacy Act). This includes making certain that anonymized data cannot be used to identify individuals and that individuals have the right to request that their personal data be deleted (ICO, 2023).

For example, the General Data Protection Regulation (GDPR) in the European Union requires organizations to anonymize personal data before using it for research, statistical analysis, or other purposes. The purpose of this regulation is to protect the fundamental right to privacy and to prevent the misuse of personal data (Directive 95/46/EC, 2018). Similarly, the Health Insurance Portability and Accountability Act (HIPAA) in the United States requires healthcare organizations to de-identify protected health information (PHI) before using it for research or other purposes (HIPAA, 2003).

Furthermore, many other countries have similar data protection laws, such as Singapore's Personal Data Protection Act, Australia's Privacy Act, and Canada's Personal Information Protection and Electronic Documents Act (PIPEDA). All of these laws require organizations to protect personal data by implementing data anonymization techniques.

5.Results

In today's digital world, data is a valuable asset for organizations, and protecting it has become a top priority. By anonymizing data, organizations are able to remove personally identifiable information from datasets, making it more difficult for unauthorized parties to identify individuals in the data. However, anonymization alone may not be enough to protect data from cyber threats. To ensure the security and confidentiality of anonymized data, organizations must implement additional technical and organisational measures. As mentioned briefly in the introduction, organizations can implement several cybersecurity measures in conjunction with data anonymization to protect data. The most effective strategy is to combine their common uses.

A real-world example of the risks associated with inadequate cybersecurity measures in anonymization is the 2018 MyFitnessPal data breach. In this incident, the personal information, including usernames, email addresses, and hashed passwords, of approximately 150 million users was stolen. Although the passwords were hashed, the attackers were able to access sensitive user data by using brute force methods due to a lack of appropriate cybersecurity measures (MyFitnessPal data breach, 2018).

This incident emphasizes the importance of strong cybersecurity measures in the anonymization process to prevent unauthorized access and potential re-identification. To ensure the security of anonymized data, organizations must implement a combination of technical and procedural safeguards, such as access controls, encryption, and regular security audits. Furthermore, employee training and awareness programs can help to foster a cybersecurity culture within an organization while reducing the risk of human error or insider threats (NIST, 2017).

The best way to assimilate the multi-model approach is to follow the steps to prevent potential threats through scenarios. We can have an empowered approach if we follow this process sequentially and include data anonymization processes.

In 2018, the National Institutes of Health (NIH) launched a new initiative called the All of Us Research Program. This program aims to collect health data from one million participants to help advance precision medicine. To protect participants' privacy, the NIH uses a combination of de-identification techniques, including removing direct identifiers such as names and addresses and modifying or removing certain data elements to further protect privacy (National Institutes of Health, 2022).

Below, we analyzed this healthcare sector scenario, where they are using de-identification techniques to protect patient privacy while allowing researchers to analyze medical data. We then integrated these stages with a security-centric approach and, finally, developed a hypothetical scenario to demonstrate the stages of the proposed security-focused multi-model data anonymization method.

- **Data Inventory and Classification:** The healthcare organization needs to first identify all the types of data it has collected and classify them based on the level of sensitivity. This process helps the organization to identify all the data that has been collected, including sensitive information, and determine the appropriate anonymization techniques to use for each type of data.

- **Data De-Identification:** The organization should employ appropriate de-identification techniques, such as removing personal identifiers such as names, addresses, and social security numbers or replacing them with pseudonyms. To further reduce the risk of re-identification, the organization should consider techniques such as data generalization or suppression. This stage involves applying appropriate de-identification techniques to remove or obscure personal identifiers to protect the privacy of individuals. Using pseudonyms, generalization, or suppression of data can help reduce the risk of re-identification.

- **Data Validation and Quality Control:** The organization should validate the anonymization process's effectiveness by testing the data for any potential re-identification risks. This can be accomplished through the use of statistical methods or other tools that assess the level of risk for re-identification. Quality control measures should also be implemented to ensure the de-identified data is accurate and reliable.

- **Secure Storage and Access Controls:** The de-identified data should be stored securely, with access controls in place to limit who can view or access the data. This can include physical controls such as secure storage rooms or electronic controls such as passwords or biometric authentication. By storing the data securely and implementing access controls, the organization can limit who can access the data and prevent unauthorized disclosure.

Some cyber security techniques can also be applied at each stage of the data anonymization process. Incorporating these processes into the data anonymization steps ensures that all of these steps have consistent security stages.

5.1 Introducing Security Centric Approach to the Data Anonymization Stages

- **Data Inventory and Classification:** Using Access Controls to limit who can view or access the data and implement network segmentation to isolate sensitive data from other parts of the network (Mohammady et al., 2021).

- **Data De-Identification:** Using Encryption to protect the data during the de-identification process and implement Data Masking or Tokenization techniques to further reduce the risk of re-identification (Ajayi & Adebiyi, 2014).

- **Data Validation and Quality Control:** Using Vulnerability Scanning and Penetration Testing to identify any weaknesses in the de-identified data and implement Data Loss Prevention (DLP) tools to detect any unauthorized attempts to access or distribute the data (Kim & Kim, 2012).

- **Secure Storage and Access Controls:** Using Encryption and Secure Hashing to protect the data at rest and implement Role-Based Access Controls (RBAC) or Attribute-Based Access Controls (ABAC) to limit who can access the data (Jin et al., 2012).

Examples of methods that can be incorporated into anonymization stages with the security-centric multi-model approach is shown in the illustration below:

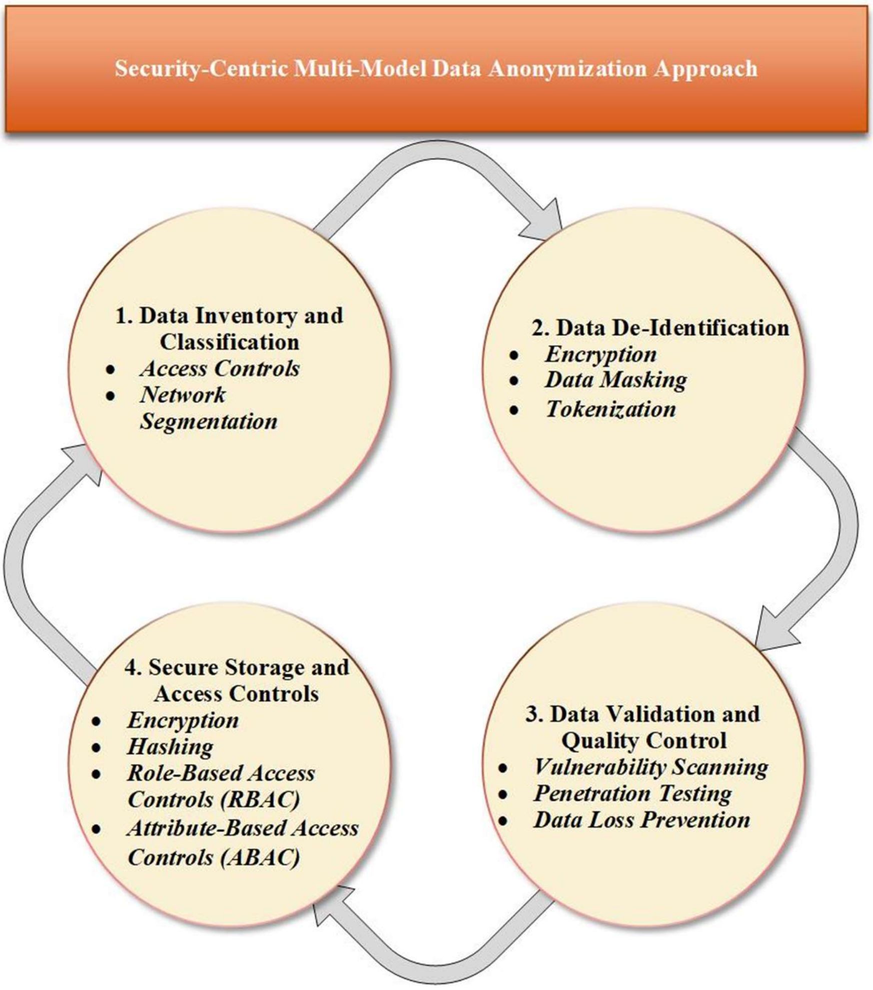

Security-Centric Multi-Model Data Anonymization Approach
1. Data Inventory and Classification
• Access Controls
• Network Segmentation
2. Data De-Identification
• Encryption
• Data Masking
• Tokenization
3. Data Validation and Quality Control
• Vulnerability Scanning
• Penetration Testing
• Data Loss Prevention
4. Secure Storage and Access Controls
• Encryption
• Hashing
• Role-Based Access Controls (RBAC)
• Attribute-Based Access Controls (ABAC)

Figure 2. Security-Centric Multi-Model Data Anonymization Approach

In order to illustrate how well these processes can work together, we have incorporated them into the hypothetical scenario presented below. We will refer to the organization as "ABC Corporation."

6.A Hypothetical Scenario with a Security-Centric Multi-Model Approach

- Data Inventory and Classification–ABC Corporation begins by conducting a data inventory and classification to identify all data types and determine their sensitivity levels. The organization uses access controls to limit who can view or access the data and implements network segmentation to isolate sensitive data from other parts of the network. Additionally, ABC Corporation uses data loss prevention (DLP) tools to detect any unauthorized attempts to access or distribute the data.

- Data De-Identification–During the de-identification process, ABC Corporation employs encryption to safeguard the data. They implement data masking or tokenization techniques to further reduce the risk of re-identification. ABC Corporation uses techniques such as k-anonymity and differential privacy to guarantee that the data has been adequately de-identified.

- Data Validation and Quality Control–After de-identifying the data, ABC Corporation performs data validation and quality control to ensure the accuracy of the de-identified data. They use vulnerability scanning and penetration testing to identify any weaknesses in the de-identified data. Additionally, ABC Corporation uses DLP tools to detect any unauthorized attempts to access or distribute the data.

- Secure Storage and Access Controls–ABC Corporation uses encryption and secure hashing to protect the data at rest. They implement role-based access controls (RBAC) or attribute-based access controls (ABAC) to limit who can access the data. ABC Corporation also performs regular audits to ensure that access controls are working appropriately.

We specified security parameters and hypothesized that "ABC Corporation" implemented the additional security measures outlined below:

- Access Controls–ABC Corporation uses Access Controls to limit who can view or access sensitive data. For example, only authorized employees with a legitimate business need can access confidential financial data.

- Network Segmentation–To further protect sensitive data, ABC Corporation has implemented Network Segmentation. This means that sensitive data is isolated from other parts of the network, making it more difficult for cyber criminals to access.

- Encryption–ABC Corporation uses Encryption to protect data during transmission and at rest. For example, sensitive data is encrypted when transmitted between servers or devices, and when stored on servers or devices.

- Data Masking–In order to further reduce the risk of data re-identification, ABC Corporation implements Data Masking. This means that sensitive data is replaced with random or fictitious data to conceal the original data. For example, instead of storing the actual social security number of an employee, the number could be masked by replacing it with a series of random numbers and letters.

- Tokenization–Similar to Data Masking, ABC Corporation also employs Tokenization techniques to further reduce the risk of re-identification. This involves replacing sensitive data with a unique token or identifier. For example, a credit card number could be replaced with a token that is only understood by authorized parties.

- Vulnerability Scanning–ABC Corporation uses Vulnerability Scanning to identify any weaknesses in their systems or processes that could be exploited by cyber criminals. For example, a regular vulnerability scan could identify an unpatched software vulnerability that needs to be addressed.

- Penetration Testing–In addition to vulnerability scanning, ABC Corporation also employs Penetration Testing to identify any weaknesses in their defenses. This involves a simulated cyber-attack that is conducted by an authorized third party to identify vulnerabilities that could be exploited by cybercriminals.

- Data Loss Prevention (DLP)–ABC Corporation uses Data Loss Prevention (DLP) tools to detect and prevent any unauthorized attempts to access or distribute sensitive data. For example, DLP tools could be configured to monitor all outgoing emails and prevent any emails containing sensitive data from being sent to unauthorized recipients.

ABC Corporation, as previously discussed, has implemented Security-Centric Multi-Model Approach to provide additional layers of security for their sensitive data. To protect their network and endpoints from cyber threats, they use a combination of technologies such as firewalls, intrusion detection systems, and anti-virus software. For example, they employ a next-generation firewall capable of detecting and blocking malicious traffic in real time.

We argue that ABC Corporation can reduce the risk of data breaches and unauthorized access to their data by implementing these methods.

7.Discussion:

The study results highlight the importance of a comprehensive and multi-model approach to data anonymization, particularly in the context of the Enterprise Data Privacy Governance Model. The favorable aspects of this approach include the ability to balance privacy protection with data quality and utility and the potential for increased effectiveness when combining various anonymization techniques and cybersecurity measures. This is in line with the literature, which emphasizes the need for holistic data privacy management to ensure compliance with legal and regulatory requirements *(Ohm, 2009)*.

However, the study also reveals some unfavorable aspects, such as the complexity of implementing multiple data anonymization techniques and the potential weaknesses of each technique. This underscores the importance of finding the right balance between these methods to create the most effective defense mechanism. Comparing the proposed security-centric multi-model approach with existing literature, it is evident that adopting a combination of techniques such as data masking, tokenization, differential privacy, machine learning algorithms, blockchain technology, and data encryption offers a robust and holistic solution to data privacy challenges (Jiang & Torra, 2023; Tachepun & Thammaboosadee, 2020; Nyugen, 2019; Mahmood & Jusas, 2022; Kshetri, 2018).

In conclusion, the multi-model approach allows organizations to capitalize on the strengths of each technique while mitigating their weaknesses, ultimately providing a more effective solution for data privacy governance.

8.Study Limitations:

There are several limitations in this study that should be acknowledged. The study primarily focuses on a theoretical framework and hypothetical scenario, which may not fully capture the challenges that organizations face when implementing data anonymization techniques in real-world settings. Future studies could benefit from examining case studies or conducting empirical research to validate the proposed multi-model approach in practice.

9.Conclusion

Data privacy is a crucial aspect of modern organizations, especially as sensitive data plays a critical role in research, policymaking, and innovation. The security-centric enterprise data anonymization governance multi-model offers a comprehensive framework that enables institutions to manage data privacy effectively while ensuring compliance with legal and regulatory requirements. By adopting this model, organizations can create a data privacy culture that permeates all operations.

However, achieving effective data anonymization is a complex task, necessitating the use of various techniques to balance privacy protection with data quality and utility. A multi-model approach that combines methods such as data masking, tokenization, differential privacy, machine learning algorithms, blockchain technology, and data encryption can provide robust and holistic solutions to data privacy challenges. This approach allows organizations to harmonize their data privacy efforts, capitalizing on the strengths of each technique while mitigating their weaknesses. The model emphasizes the importance of assessing the organization's data privacy risk and implementing appropriate controls to mitigate these risks.

Ultimately, the combination of the security-centric enterprise data anonymization governance model and a multi-model approach to data anonymization can help organizations protect sensitive information while maintaining data utility for research and analysis. By doing so, institutions can strike the delicate balance between preserving individual privacy and leveraging data to drive innovation and growth.

References:

Ajayi, O. O., & Adebiyi, T. O. (2014). Application of Data Masking in Achieving Information Privacy. *IOSR Journal of Engineering,* 4(2), 13-21.

Cavoukian, A., & Jonas, J. (2011). Privacy by design: A framework for designing privacy into the new technologies. *Identity in the Information Society*, 4(1), 3-23. doi: 10.1007/s12394-010-0052-3

Directive 95/46/EC, (2016). Retrieved from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434

El Emam, K., Jonker, E., Arbuckle, L., & Malin, B. (2011). A systematic review of re-identification attacks on health data. *Plos One,* 6(12), 28071. https://doi.org/10.1371/journal.pone.0028071

HIPAA, (2017). De-identification standard, Retrieved from: https://www.govinfo.gov/content/pkg/CFR-2017-title45-vol1/pdf/CFR-2017-title45-vol1-sec164-514.pdf (accessed on 26 March 2023).

ICO, (2021). *How do we ensure anonymisation is effective?* https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf (accessed on 26 March 2023).

IDStrong (2022). MyFitnessPal Breach: Learn About MyFitnessPal Hack. https://www.idstrong.com/sentinel/myfitnesspal-data-breach/ (accessed on 26 March 2023.)

Jassim, H., Atan, R., Jabar, M., & Abdullah, S. (2018). Factors and model for sensitive data management and protection in information systems' decision of cloud environment. *Journal of Theoretical and Applied Information Technology,* 96, 8097–8108.

Jiang, L., & Torra, V. (2022). On the Effects of Data Protection on Multi-database Data-Driven Models, *Integrated Uncertainty in Knowledge Modelling and Decision Making,* 226–238.

Jin, X., Krishnan, R., & Sandhu, R. (2012). *A unified attribute-based access control model covering DAC, MAC and RBAC.* In Data and Applications Security and Privacy XXVI: 26th Annual IFIP WG 11.3 Conference, 26, 41-55.

Kalloniatis, C., Mouratidis, H., Vassilis, M., Islam, S., Gritzalis, S., & Kavakli, E. (2014). Towards the design of secure and privacy-oriented Information Systems in the Cloud: Identifying the major concepts. *Computer Standards & Interfaces,* 36(4), 759-775. https://doi.org/10.1016/j.csi.2013.12.010

Kim, J., & Kim, H. J. (2012). *A Study on Privacy Preserving Data Leakage Prevention System.* In Recent Progress in Data Engineering and Internet Technology, 2 191-196.

Kshetri, N. (2018). Blockchain's roles in meeting key supply chain management objectives. *International Journal of Information Management,* 39, 80-89. doi: 10.1016/j.ijinfomgt.2017.12.001

LeFevre, K., DeWitt, D. J., & Ramakrishnan, R. (2006). *Mondrian multidimensional k-anonymity.* In 22nd International conference on data engineering, 25-25.

Lu, J., & Holubová, I. (2019). Multi-Model Databases: A New Journey to Handle the Variety of Data. *ACM Computing Surveys,* 52(3).

Lubarsky, B. (2017). Re-Identification of "Anonymized" Data. *Georgetown Law Technology Review,* 202, 1-12.

Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). L-Diversity: Privacy beyond k-Anonymity. *The ACM Transactions on Knowledge Discovery from Data.* 1(1), 3.

Mackey, E., Elliot, M., & O'Hara, K. (2016). *The anonymisation decision-making framework.* (1st Edition). UK Anonymization Network.

Mahmood, Z., & Jusas, V. (2022). Blockchain-Enabled: Multi-Layered Security Federated Learning Platform for Preserving Data Privacy. *Electronics,* 11(10), 1624. http://dx.doi.org/10.3390/electronics11101624

Mourby, M., Mackey, E., Elliot, M., Gowans, H., Wallace, S. E., Bell, J., … Kaye, J. (2018). Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research. *UK. Computer Law & Security Review,* 34(2), 222–233. doi:10.1016/j.clsr.2018.01.002

Nguyen, A. (2022). Understanding Differential Privacy. Retrieved from: https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

NIH, (2022). Achieving the Principles through a Precision Medicine Initiative Data Security Policy Framework, All of Us Research Program. Retrieved from: https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-data-security-policy-principles-and-framework-overview/achieving-principles-through-precision-medicine-initiative-data-security-policy-framework

NIST, (2014). Framework Version 1.0, Retrieved from: https://www.nist.gov/cyberframework/draft-version-11

Ohm, P. (2009). Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57, 1701.

Oqaily, M., Jarraya, Y., Mohammady, M., Majumdar, S., Pourzandi, M., Wang, L., & Debbabi, M. (2021). SegGuard: Segmentation-Based Anonymization of Network Data in Clouds for Privacy-Preserving Security Auditing. *The IEEE Transactions on Dependable and Secure Computing*, 18(5), 2486–2505. doi:10.1109/TDSC.2019.2957488

Samarati, P., & Sweeney, L. (1998). *Generalizing Data to Provide Anonymity When Disclosing Information (Abstract).* Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 188. doi:10.1145/275487.275508

Statista, (2023). Retrieved from: Number of data breaches and victims U.S. Retrieved from: https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/

Sweeney, L. (2002). Achieving k-Anonymity Privacy Protection Using Generalization and Suppression. *The International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 10, 571-588.

Tachepun, C., & Thammaboosadee, S. (2020, July). *A Data masking guideline for optimizing insights and privacy under GDPR compliance.* In Proceedings of the 11th international conference on advances in information technology, 1-9.

Xu, J., Wang, W., Pei, J., Wang, X., Shi, B., & Fu, A. W. C. (2006, August). *Utility-based anonymization using local recording.* In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 785-790.

Zhou, B., & Pei, J. (2011). The k-anonymity and l-diversity approaches for privacy preservation in social networks against neighborhood attacks. *Knowledge and information systems*, 28(1), 47-77.