Journal of Cybersecurity and Privacy

MDPI

Article

Data Privacy and Ethical Considerations in Database Management

Eduardo Pina [1], José Ramos [1], Henrique Jorge [1], Paulo Váz [2], José Silva [2], Cristina Wanzeller [2], Maryam Abbasi [3] and Pedro Martins [2,*]

[1] Department of Informatics, Polytechnic of Viseu, 3504-510 Viseu, Portugal; pv27228@alunos.estgv.ipv.pt (E.P.); pv21027@alunos.estgv.ipv.pt (H.J.)
[2] Research Center in Digital Services, Polytechnic of Viseu, 3504-510 Viseu, Portugal; paulovaz@estgv.ipv.pt (P.V.); jsilva@estgv.ipv.pt (J.S.); cwanzeller@estgv.ipv.pt (C.W.)
[3] Applied Research Institute, Polytechnic of Coimbra, 3045-093 Coimbra, Portugal; maryam.abbasi@ipc.pt
* Correspondence: pedromom@estgv.ipv.pt

Abstract: Data privacy and ethical considerations ensure the security of databases by respecting individual rights while upholding ethical considerations when collecting, managing, and using information. Nowadays, despite having regulations that help to protect citizens and organizations, we have been presented with thousands of instances of data breaches, unauthorized access, and misuse of data related to such individuals and organizations. In this paper, we propose ethical considerations and best practices associated with critical data and the role of the database administrator who helps protect data. First, we suggest best practices for database administrators regarding data minimization, anonymization, pseudonymization and encryption, access controls, data retention guidelines, and stakeholder communication. Then, we present a case study that illustrates the application of these ethical implementations and best practices in a real-world scenario, showing the approach in action and the benefits of privacy. Finally, the study highlights the importance of a comprehensive approach to deal with data protection challenges and provides valuable insights for future research and developments in this field.

Keywords: data privacy; ethics; database management; best practices

Citation: Pina, E.; Ramo, J.; Jorge, H.; Váz, P.; Silva, J.; Wanzeller, C.; Abbasi, M.; Martins, P. Data Privacy and Ethical Considerations in Database Management. *J. Cybersecur. Priv.* **2024**, *4*, 494–517. https://doi.org/10.3390/jcp4030024

Academic Editor: Martin Gilje Jaatun

Received: 28 May 2024
Revised: 14 July 2024
Accepted: 26 July 2024
Published: 29 July 2024

1. Introduction

Data privacy and the rise of ethical challenges continue to undergo a detailed analysis as part of the digital revolution. Safeguarding data privacy constitutes a fundamental right, often overlooked amid the exchange of data transfer for commercial and scientific purposes [1].

The ethical considerations in data management extend beyond mere privacy concerns. They encompass a broader spectrum of responsibilities, including the fair and transparent use of data, ensuring data integrity, and respecting the rights of data subjects. In this context, ethics in data management can be defined as the principles and practices that guide the responsible collection, processing, storage, and use of data, with a focus on protecting individual rights and maintaining public trust.

Requiring commodities to handle data securely and ethically allows this initiative to protect individual's data control rights and builds user trust. Furthermore, while protecting individual rights, it ensures compliance with strict regulations to mitigate risks such as unauthorized access, misuse, or sensitive data breaches. As a result, understanding the complexities of data privacy is critical for advancing research, technology, and policy-making areas in the digital age [2]. The combination of all these issues with the ongoing advancements in technology and concerns about the improper use of personal data by governments and corporations compelled the application of laws to clarify data privacy rights and ensure an appropriate worldwide level of protection for personal data [3].

This paper primarily draws its requirements for ethical and privacy considerations from two key regulations: the General Data Protection Regulation (GDPR) [4] and the

Health Insurance Portability and Accountability Act (HIPAA) [5]. These regulations were chosen due to their comprehensive nature and wide-ranging impact on data management practices globally. The GDPR, a regulation (not a directive) in EU law, provides a robust framework for data protection and privacy in the European Union and the European Economic Area. HIPAA, on the other hand, offers specific guidelines for protecting sensitive patient health information in the United States. Together, these regulations provide a solid foundation for discussing data privacy and ethical considerations in database management.

It is important to note that, throughout this paper, we will use the term "data subject" to refer to individuals whose personal data is being processed, aligning with the terminology used in the GDPR. This choice reflects our focus on regulatory compliance and helps in precisely scoping the requirements discussed in this paper.

There are regulations such as the aforementioned General Data Protection Regulation (GDPR) [4] and Health Insurance Portability and Accountability Act (HIPAA) [5] that assist organizations in safeguarding personal data under their jurisdiction, ensuring that organizations are informed and able to react to such an occurrence. In addition, HIPAA and GDPR impact database systems significantly, which affect their architecture, security protocols, and user access controls [6]. Adherence to these regulations requires incorporating features that aid in following their requirements, like fields for consent statuses and attributes related to data classification in database schemas [7]. Given these regulations, access to data is carefully controlled, following the privilege principle to grant access to only authorized users. Security measures have been tightened to protect sensitive data, needing robust encryption methods, intrusion detection systems, and strict access controls. These are some characteristics implemented by these regulations [8].

Database administrators (DBAs) are vital in ensuring stored data security, integrity, and confidentiality. They are essential in ensuring ethically sound data practices within corporate environments. Their principal responsibility lies in implementing robust data security measures to safeguard databases against potential threats posed by unauthorized access and malicious actors [9] while taking into consideration database design, security protocols, and user access permissions to comply with HIPAA and GDPR regulations [10].

A critical ethical consideration in database management is the right of data subjects to request the removal of their data, often referred to as the "right to be forgotten" in the GDPR. Database designs and management practices must incorporate mechanisms to efficiently and completely remove an individual's data upon request, while maintaining the integrity of the remaining data and complying with any legal retention requirements.

This paper presents a practical application of data privacy principles in managing a patient health dataset that mimics real-world healthcare data. The dataset was acquired from Kaggle [11], a popular open-source platform that presents datasets, machine learning competitions, tools, and tutorials for data science enthusiasts and additional areas.

The results obtained from implementing data minimization, anonymization, pseudonymization, encryption, and access control practices in the "Healthcare Dataset" demonstrate significant improvements in the security and privacy of stored health information, although with a slight performance impact. Data minimization reduces the stored sensitive information, lowering the risk of exposure and privacy breaches. Anonymization and pseudonymization add a layer of protection while reducing data granularity, making detailed analysis challenging. Encryption using the Advanced Encryption Standard algorithm, facilitated by the *pgcrypto* extension in PostgreSQL, effectively safeguard sensitive data from unauthorized access, although at the cost of increased processing time. Lastly, access control measures enhance system security by carefully assigning permissions to database users.

In addition to encryption, this paper will also discuss hashing techniques, which are crucial for verifying data integrity. We will explore various hashing algorithms and their applications in database management, complementing our discussion on encryption methods.

This paper is structured as follows. Section 2 provides a literature review on works related to database management, data privacy, and ethical considerations. In Section 3,

we introduce specific techniques and rules that database administrators should follow regarding ethical issues such as minimization, anonymization, pseudonymization and data encryption, access controls, and transparent communication with stakeholders. Section 4 presents a real-world database scenario that shows how to implement ethical implementations and best practices. Section 5 discusses the results and analyses obtained by applying ethical considerations and best practices. In Section 6, we provide recommendations for more research in this area, critically evaluate the effects and implications of our findings, and summarize the study's findings.

Throughout this paper, we aim to provide a comprehensive overview of data privacy and ethical considerations in database management, addressing not only technical aspects but also the broader ethical implications of data handling practices.

2. Literature Review

This section presents works related to data privacy and ethical concerns in database management.

In one paper [12], the authors present a tool used in PostgreSQL to mask and anonymize the information of a database. They show techniques like masking parts of the text commonly applied to credit card numbers, where only the last four digits remain visible; date aging that substitutes the value with a range of ages; and nullifying columns where the actual values are swapped for a null value. This technique shows an example of how to help database managers protect critical user information to prevent data infringements.

Another paper [13] presents an analysis of the performance of encryption tools over three databases, where one of the databases mentioned is PostgreSQL. In this database, the authors applied a tool called pgcrypto. This tool provides several functions, such as creating password hashes, Pretty Good Privacy (PGP) encryption functions, and random hash functions. The authors applied these security measures and evaluated the performance of databases given different workloads. The author states that PostgreSQL provides good security over data protection, although its performance may be slightly impacted.

The authors proposed a study on privacy methods and anonymization techniques to protect sensitive data on databases to prevent attacks and data recognition to evaluate its performance [14]. Some methods applied were k-anonymity, l-diversity, suppression, generalization, anatomization, and slicing. The authors deduce that they must lose some information to protect data privacy and prevent data recognition, even though attacks can still be performed.

The authors of [15] present a study that analyses an anonymization process of databases of banking applications while applying anonymization methodologies to improve its security. The authors found some challenges, given that the anonymization process can be very complex and time consuming, mainly when dealing with vast amounts of data. In addition, they explain the different contexts, whether organizational, functional/business, or technical, when applying anonymization techniques.

Article [16] addresses the possible risks and challenges related to data minimization, which consists of limiting the amount of personal information collected, stored, and processed to only what is necessary to achieve a specific objective. Despite being commonly recommended as an excellent privacy measure, the article points out that data minimization can have unexpected and potentially harmful consequences.

In one paper, the researchers demonstrated that, by carefully minimizing the data collected about a student's daily activities, they could still obtain valuable insights for personalized learning while effectively protecting the student's privacy. This application shows how data minimization can be used creatively in specific contexts to combine the need for data-driven decision making with robust privacy protections [17].

Other authors published a study outlining a methodology for pseudonymizing medical data [18]. This involved hashing and encrypting patient identifiers and securely storing the encryption key, allowing for data analysis while preserving patient confidentiality.

A study carried out in [19] highlights the importance of choosing algorithms when comparing data encryption performance. The research concluded that, while solid encryption is essential, the algorithm also needs to ensure adequate processing to avoid possible negative impacts on the performance of the selected database.

In the context of ethical considerations in data management, the Royal Society's report on "Privacy Enhancing Technologies" [20] provides valuable insights into the intersection of technology and privacy. The report discusses various technologies and methodologies that can be employed to enhance privacy in data-intensive applications, which is particularly relevant to our discussion on ethical database management.

Recent developments in data governance and protection regulations, such as the EU's Data Act [21], Data Governance Act [22], Digital Services Act (DSA) [23], and the proposed EU AI Act [24], are also worth considering for future research. These regulations introduce new requirements and considerations for data management and use, particularly in the context of AI and digital services.

Addressing the ethical implications of data management, a study by [25] explores the ethical challenges posed by big data. The authors discuss issues such as privacy, anonymity, transparency, and trust in the context of large-scale data collection and analysis. This work provides a broader perspective on the ethical considerations that should inform database management practices.

In response to the growing concern over the right to be forgotten, ref. [26] discusses the technical and legal challenges of implementing this right in database systems. The paper explores the tensions between data retention, the right to privacy, and the practical implications for database design and management.

Regarding the specific requirements derived from GDPR and HIPAA, ref. [27] provides a comprehensive analysis of the GDPR's impact on data management practices. Similarly, ref. [28] offers insights into the implications of HIPAA for healthcare data management. These works help in understanding the specific requirements that shape ethical and privacy considerations in database management.

The work performed on these papers may differ from our paper but will serve as a guide to conducting our study.

3. Best Practices

This section presents ethical best practices for database administrators, such as data retention policies, data minimization, data anonymization and pseudonymization, access controls, data encryption, data hashing, and transparent communication with stakeholders.

3.1. Data Retention Policies

Data retention policies have a vital role in upholding data security and confidentiality. Their implementation and upkeep carry considerable weight for businesses, particularly those with data-centric models like research service providers [29]. Without indefinite data retention, there is a risk of losing collective memory, leading to a misleading perception of privacy and freedom [**?**].

Therefore, some of the best methods to develop data retention policies are the following:

- Data assessment: Identify and classify the types of data the organization stores, considering the level of risk to which they are subject. Personal data, such as Personally Identifiable Information (PII), should be prioritized and managed rigorously.
- Clear conservation periods: Collaboration with data owners to establish and validate appropriate retention periods for diverse data categories, balancing legal and operational mandates while safeguarding confidentiality.
- Deletion processes: Implement efficient strategies for purging obsolete or irrelevant data within specified retention timelines, mitigating data accumulation and minimizing the risk of privacy breaches or unauthorized disclosures.

- Restricted access control: Ensure strict control over access to sensitive data, permitting only the authorized personnel necessary for organizational functions while maintaining comprehensive monitoring and recording mechanisms to detect and address any potential unauthorized breaches.
- Continuous assessment and improvement: View data retention policies as dynamic, requiring ongoing assessment and enhancement, with data controllers staying up to date on evolving legal frameworks, security protocols, and technological advancements to ensure compliance and maximize effectiveness.

While following the above methods, some key principles are needed to follow to build data retention policies:

- Have a precise, legal, and legitimate purpose;
- Establish a precise and time-limited retention period;
- Data must be relevant and necessary;
- Data must be stored for the shortest possible time;
- Data is safe and secure.

As with many legal policies, they present several implications regarding the implementation of these policies:

- Data volume management: The surge in data volume needs efficient storage, processing, and deletion approaches to maintain critical organizational data while reducing the threat of excessive accumulation.
- Difficulty in data classification: Ensuring accuracy and consistency in data classification is essential for successfully applying retention policies. However, organizations need help in accurately classifying data.
- Ongoing regulatory changes: Data controllers must stay updated with changes to relevant laws and regulations, as well as security and privacy best practices and technologies.

Nonetheless, implementing data retention policies is challenging. Data managers face balancing operational needs with strict compliance requirements [29].

3.2. Ethical Considerations in Data Retention

When implementing data retention policies, several ethical considerations must be taken into account:

- Right to be forgotten: As per GDPR Article 17, data subjects have the right to request the erasure of their personal data. Database designs must incorporate mechanisms to efficiently and completely remove an individual's data upon request [31].
- Purpose limitation: Data should only be collected and retained for specified, explicit, and legitimate purposes, as outlined in GDPR Article 5(1)(b) [32].
- Data minimization: Organizations should limit the collection and retention of personal data to what is necessary for the specified purposes, in line with GDPR Article 5(1)(c) [33].
- Storage limitation: Personal data should be kept in a form which permits the identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed, as per GDPR Article 5(1)(e) [34].

These ethical considerations ensure that data retention policies not only comply with legal requirements but also respect the rights and privacy of data subjects.

3.3. Data Minimization

Regarding database security, a practical approach to safeguarding sensitive information involves reducing collected and processed data [8]. As such, data collection is appropriate and minimizes the risks of unauthorized access, data breach, and disorderly use, which aligns with the general principle of collecting the minimum data necessary to achieve a specific purpose like GDPR.

For DBAs, implementing data minimization requires careful collaboration with other stakeholders to define essential data elements and formulate data guidelines [35]. This also applies strict controls over data processing activities to prevent the intentional or accidental dissemination of sensitive information. By reducing the overall data footprint and limiting access to only authorized personnel, DBAs can significantly improve the security view of their databases [36].

Tables 1 and 2 show an example of how data minimization works in the context of students' data:

Table 1. Student data information before data minimization.

Student ID	Name	Date of Birth	Gender	Address	Email	Phone	Course	Year	GPA	Guardian1	Guardian1_Email	Guardian_Phone
001	John Smith	20 May 1998	Male	123 Main St, Anytown	john.smith@example.com	555-1234	Computer Science	2nd	3.5	Mary Smith	mary.smith@example.com	555-1111
002	Emily Johnson	15 September 1999	Female	456 Elm St, Anycity	emily.j@example.com	555-5678	Biology	3rd	3.8	Lisa Johnson	lisa.johnson@example.com	555-3333
003	David Lee	10 February 2000	Male	789 Oak St, Anyville	david.lee@example.com	555-9876	History	1st	3.2	Linda Lee	linda.lee@example.com	555-5555
004	Sarah Brown	30 November 1997	Female	321 Pine St, Anysuburb	sarah.b@example.com	555-4321	Mathematics	4th	3.9	Karen Brown	karen.brown@example.com	555-7777

Table 2. Student data information after data minimization.

Student ID	Name	Course	Year	GPA
001	John Smith	Computer Science	2nd	3.5
002	Emily Johnson	Biology	3rd	3.8
003	David Lee	History	1st	3.2
004	Sarah Brown	Mathematics	4th	3.9

This example illustrates how data minimization works given the immense information shown in Table 1. By carefully minimizing the data collected about a student's personal information, one can select important information without disclosing other information to protect the student's privacy.

3.4. Ethical Implications of Data Minimization

While data minimization is generally considered a good privacy practice, it is important to consider its potential ethical implications:

- Loss of context: Excessive data minimization might lead to a loss of important contextual information, potentially affecting the quality of data analysis or decision-making processes [16].
- Bias introduction: If not carefully implemented, data minimization could inadvertently introduce bias by selectively removing certain types of data [37].
- Balancing utility and privacy: There is an ongoing ethical challenge in finding the right balance between minimizing data for privacy protection and retaining enough data for meaningful analysis and use [17].

DBAs and organizations must carefully consider these ethical implications when implementing data minimization strategies to ensure they protect privacy without compromising the utility and fairness of their data practices.

3.5. Anonymization and Pseudonymization of Data

Anonymization and pseudonymization methods are critical in safeguarding confidential data [38]. These techniques replace directly identifiable information with non-identifiable alternatives, significantly reducing the risk of individuals being re-identified within the data.

Anonymization involves removing or concealing all Personally Identifiable Information (PII) from a database. Techniques such as hashing, tokenization, and generalization replace specific data with non-identifying codes or modify it to remove specificity [39]. For example, in Table 3, the application of anonymization techniques overwrites the name of the subject with fictitious personal information, as illustrated in Table 4.

Table 3. Subject names before anonymization.

Subject Name	Disabilities
Alice Smith	Vision impairment
Bob Johnson	None
Dave Doe	Deaf or hard of hearing
Eve Jackson	None
Grace Chan	Mental health

Table 4. Subject names after anonymization.

Subject Name	Disabilities
Rgwko Bkgoeoe	Vision impairment
Ekgogks Gbkoskfso	None
Gege Wuiwgb	Deaf or hard of hearing
Boxoe Bolw	None
PogiW Bikswekm	Mental health

Pseudonymization is another approach to protect privacy, particularly in healthcare settings [40]. For instance, patient records in a healthcare database may be pseudonymized by hashing the patient's name and date of birth and storing the hash and encrypted data. Pseudonymized data cannot be directly linked to a specific individual without additional information. Table 5 shows the student name before pseudonymization and Table 6 shows the student name after pseudonymization.

Table 5. Student names before pseudonymization.

Student Name	Grades	Study Hours
Pedro Silva	10.0	55
Maria Eduarda	7.0	40
Lucas Lima	9.0	42
Carlos Melo	8.5	50
Beatriz Gomes	6.0	30

Table 6. Student names after pseudonymization.

Student Name	Grades	Study Hours
12576	10.0	55
34981	7.0	40
21067	9.0	42
48007	8.5	50
36902	6.0	30

There are other techniques used in the anonymization process to protect data [39] that include the following:

- Remove Attributes (Suppression): This method eliminates attributes from the dataset that could help identify individuals. It should be used when an attribute is irrelevant or unnecessary for analysis or when anonymization is unfeasible by any other means.
- Character Replacement (Masking): This technique involves using neutral characters, such as the "*", to hide personal and vital information. This replacement can partially hide a text or attribute, which can be sufficient to anonymize the data.
- Scrambling/Shuffling: This technique involves randomly mixing or rearranging the data while retaining the values of the original attributes in the dataset. It can analyze individual attributes independently without requiring their correlation with others.

- Data Noise (Perturbation): This technique, also known as data perturbation, consists of minor modifications to the dataset's attributes to make them less precise and remove possible identifications. It is crucial to understand the level of noise that should be applied to the data without compromising an individual's analysis or privacy.
- Generalization: This technique involves modifying attributes by changing their scale or order of magnitude to provide an overview of each attribute.
- Aggregation: This method condenses data into summarized versions with fewer attributes through standardization and grouping similar data. Aggregation differs from generalization in that it actively alters the data, as opposed to the basic setting applied to each attribute in generalization.

The anonymization and pseudonymization techniques are essential for granting regulations such as GDPR and HIPAA. They enable organizations to lawfully process and share personal data for various purposes while mitigating the risk of unauthorized access and misuse.

3.6. Ethical Considerations in Anonymization and Pseudonymization

While anonymization and pseudonymization are crucial for protecting privacy, they also raise several ethical considerations:

- Re-identification risk: Even with these techniques, there is always a risk of re-identification, especially with advances in data analytics. Ethical database management requires ongoing assessment of these risks [18].
- Data utility vs. privacy: There is often a trade-off between the level of anonymization and the usefulness of the data. Striking the right balance is an ethical challenge [41].
- Informed consent: It is crucial to consider whether data subjects have given informed consent for their data to be anonymized or pseudonymized, especially if the original purpose of data collection changes [42].
- Transparency: Organizations should be transparent about their anonymization and pseudonymization processes to maintain trust with data subjects [43].

These ethical considerations highlight the need for a thoughtful and balanced approach to implementing anonymization and pseudonymization techniques in database management.

3.7. Data Encryption

In database security, data encryption is essential to protect sensitive information. DBAs can establish efficient protection against unauthorized and potential data access by selectively encrypting critical data and limiting access to encryption keys [35].

The efficacy of data encryption hinges on adherence to various best practices. First, meticulously evaluating the data types that justify encryption is essential. Sensitive data, including Personally Identifiable Information (PII), Protected Health Information (PHI), and proprietary data must be prioritized. Additionally, DBAs must opt for appropriate encryption algorithms that balance security strength and practical performance considerations. Regarding database encryption, data protection can be implemented across multiple levels, including cell, column, tablespace, and file levels.

There are two types of Data Encryption Techniques:

- Symmetric database encryption uses the same secret key for data encryption and decryption as illustrated in Figure 1. It is a simple and effective solution at a low cost. The main problem with this method is keeping the key secret and sharing it securely among authorized users. Examples of symmetric encryption algorithms include 3DES, AES, DES, QUAD, and RC4.

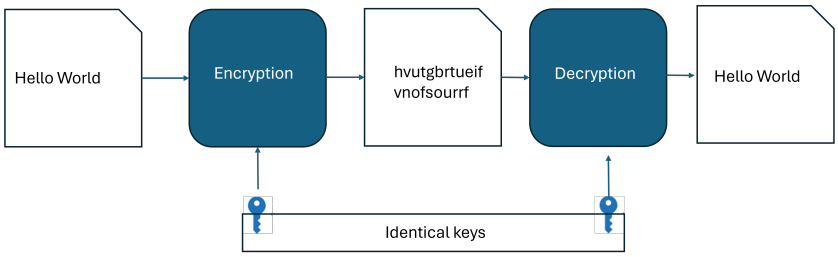

Figure 1. Symmetric encryption.

- Asymmetric database encryption, also known as public-key encryption, encrypts and decrypts data using public and private keys as shown in Figure 2. A more secure method for protecting sensitive data is to use a public key, which can only be decrypted given an associated private key. By separating public and private keys, asymmetric database encryption unravels the problem of managing key distribution. Examples of asymmetric encryption algorithms include Diffie–Hellman, ECC, El Gamal, DSA and RSA.

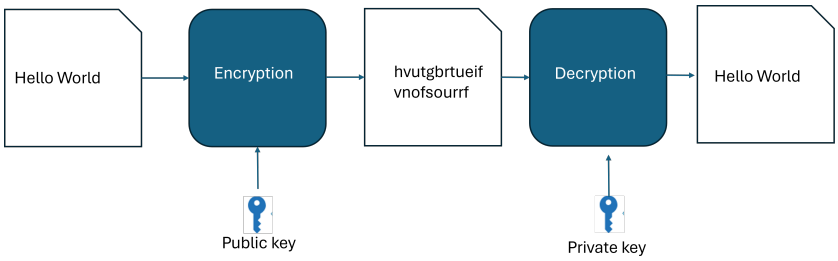

Figure 2. Asymmetric encryption.

Data encryption is essential for database security, ensuring confidentiality and appropriately maintaining the integrity and availability of information. The Data Encryption Standard (DES) has been fundamental to information security among the several symmetric encryption algorithms. By transforming plain text data into cypher text using a shared secret key, these algorithms ensure that the encrypted data will remain indecipherable without the key in case of unauthorized access to the database [44]. DES is no longer considered safe for new systems or applications requiring high security. The Advanced Encryption Standard (AES) is a symmetric key encryption algorithm that replaces DES and is widely used in modern systems and applications that require a high level of security [45].

3.8. Data Hashing

Hashing is a fundamental cybersecurity and data management technique that transforms data into a hash value, a fixed-length sequence of characters. This procedure ensures that the data is represented by a distinct and different identifier called a hash. Hashing is widely used because it allows for the verification of data integrity and the protection of confidential information without exposing the original data.

Many programs can transform the text into a hash string. Common hashing algorithms include the following:

- MD5: A hash function compresses any length of data processes or messages into a 128 bit hash value [46], where the same input will always produce the same output and, as it is unidirectional, the process cannot be reversed. Figure 3, from [47], demonstrates how the MD5 algorithm is used.
  Even though it can be widely used for certificate authentication, it has vulnerabilities regarding hash collision [48]. Due to this, it is no longer acceptable within secure hash functions where collision resistance is required, such as in the digital signatures stated by [49]. This function can be used or is preferred due to its lower computational requirements than recent Secure Hash Strings.

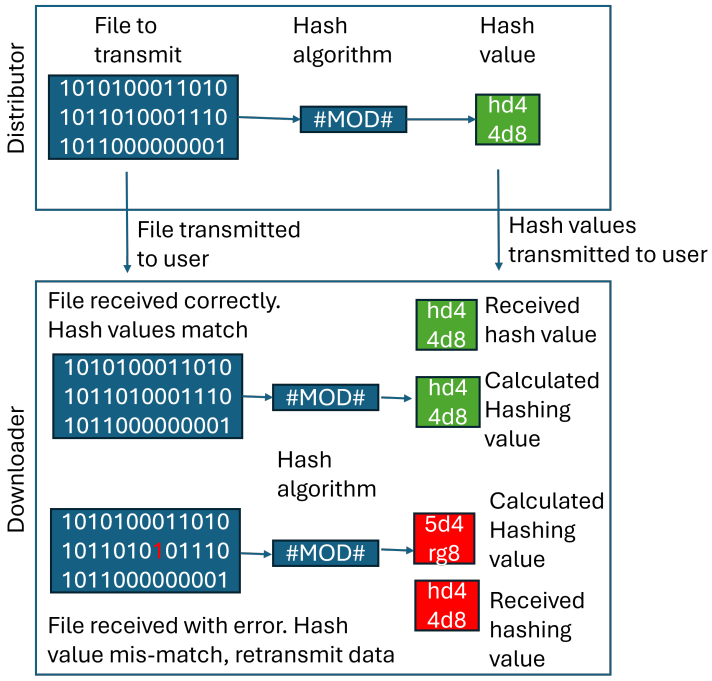

Figure 3. MD5 function.

- SHA-256: The hash function successor of SHA-1 is one of the most robust hash functions available [50] and is a highly secure one-way hash function that makes it impossible to recover the original message from the hash value [51]. The SHA256 architecture describes the steps of message processing, including message input, padding, expansion, scheduling, and compression, to produce a 256-bit hash value. Figure 4 demonstrates the different stages, including Message Padding for message length adjustment, Message Expansion to determine the original message length before padding, Message Scheduler to schedule message block processing, and Message Compression to generate a 256-bit message digest after 64 rounds [51].

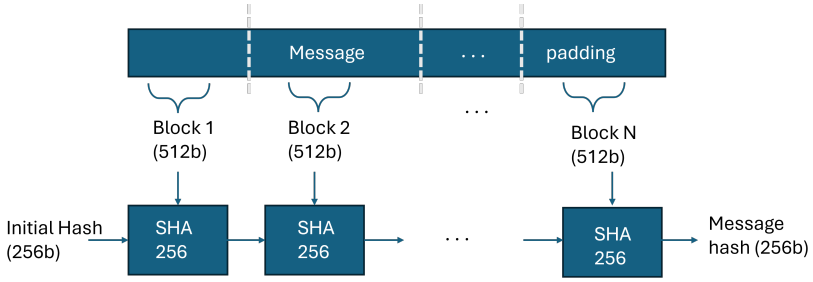

Figure 4. SHA-256 architecture.

In addition to MD5 and SHA-256, other hashing algorithms such as SHA-3, Blake2, and Whirlpool are also employed in various applications, offering a range of security and performance features to suit different cybersecurity and data management needs. These hashing algorithms ensure that multiple options are available, each with specific requirements for data integrity verification and secure information protection, enhancing the overall security of various systems.

3.9. Access Controls

In database security, establishing access controls is paramount since it will likely reduce the likelihood of security incidents [52]. This process aligns with gathering necessary data and restricting access to prevent unauthorized disclosures, data breaches, and other security breaches. For database administrators (DBAs), integrating access control principles into the

core of software systems is crucial for maintaining information integrity, confidentiality, and availability [53].

In addition to technical controls, developing clear guidelines and procedures for access control is essential. These guidelines should specify who can access specific data, under what circumstances, and for what purpose.

3.10. Ethical Considerations in Access Control

Implementing access controls raises several ethical considerations:

- Privacy vs. Utility: Balancing the need for data protection with the need for data access to perform necessary functions is an ongoing ethical challenge [54].
- Transparency: Organizations should be transparent about their access control policies to maintain trust with data subjects and stakeholders [55].
- Non-discrimination: Access control measures should be implemented in a way that does not unfairly discriminate against certain groups or individuals [56].
- Accountability: There should be mechanisms in place to ensure that those who access data are accountable for their actions [57].

These ethical considerations underscore the importance of the thoughtful and fair implementation of access controls in database management.

3.11. Transparent Communication with Stakeholders

Effective and transparent communication with target audiences and other stakeholders is crucial for addressing the challenges and opportunities of data management [58]. Communication is a powerful tool that explains limitations and potential risks, followed by decision making. It is essential to find the right balance in communication to avoid excesses that cause information fatigue and lack of communication that leads to misunderstandings and mistrust.

Regarding ethical considerations in data management, it is essential to deal with issues related to data privacy [59]. Good data management practices must follow evolving privacy regulations, ensuring that sensitive information is collected, stored, and used following legal guidelines [60]. By providing clear information about their data practices and allowing people to control their personal information, organizations can show their commitment to ethics.

To communicate efficiently with stakeholders, organizations need to implement best practices for communicating with these stakeholders [61]. This encompasses a four-step process: planning, implementation, monitoring, and evaluation [61].

Ideal practices for transparent communication are as follows:

- Establish clear objectives and roles: Allows all stakeholders to understand the goals and their affiliated roles. This transparency contributes to clarity and establishes the alignment of expectations from the beginning.
- Two-way communication: Transparency involves maintaining open communication channels with stakeholders regarding projects or initiative progress. Regular updates for stakeholders to make inquiries and feedback encourage a sense of inclusion and mutual dedication.
- Sharing positive and negative information: Demonstrating sincerity builds credibility and shows that the organization is not concealing information.
- Use clear, concise language: Avoid technical terminology and overly complex terms that non-technical stakeholders may find challenging to comprehend. The goal is to ensure that information is accessible and easily understood by all relevant stakeholders.
- Schedule regular communication: Set a periodic calendar of stakeholder meetings. This may involve weekly status updates, monthly progress reports, or ad hoc meetings as needed. Keeping all parties informed and aligned is facilitated by consistent communication.

3.12. Ethical Implications of Transparent Communication

Transparent communication in data management carries several ethical implications:

- Trust building: Transparent communication helps build trust between organizations and their stakeholders, including data subjects [62].
- Informed consent: Clear communication is crucial for ensuring that data subjects can give truly informed consent for the use of their data [63].
- Accountability: Transparency in communication promotes accountability in data management practices [64].
- Empowerment: By providing clear information about data practices, organizations empower individuals to make informed decisions about their personal data [65].

These ethical implications highlight the importance of prioritizing transparent communication in all aspects of data management.

4. Case Study

We used a dataset that mimics an example of a real-world health dataset, given that authentic patient healthcare data cannot be used due to privacy concerns. The dataset "Healthcare Dataset" was acquired from Kaggle, a popular data science competition platform and online community for data scientists and machine learning professionals. This dataset contains patient-specific information, making it an essential target for privacy-related concerns.

To manage the dataset data, we used pgAdmin4 v8.2, which is a database administration tool for PostgreSQL.

The dataset contains the following columns:

- Name: This column indicates the patient's name associated with the healthcare record.
- Age: This column indicates the age of the patient at the time of admission, expressed in years.
- Gender: This column indicates the gender of the patient, either "Male" or "Female".
- Blood Type: The patient's blood type, which can be one of the common blood types (e.g., "A+", "O−", etc.).
- Medical Condition: This column specifies the primary medical condition or diagnosis associated with the patient, such as "Diabetes", "Hypertension", "Asthma", etc.
- Date of Admission: This column specifies the date the patient was admitted to the healthcare facility.
- Doctor: This column indicates the name of the doctor responsible for the patient's care during their admission.
- Hospital: This column identifies the healthcare facility or hospital where the patient was admitted.
- Insurance Provider: This column indicates the patient's insurance provider, which can be one of several options, including "Aetna", "Blue Cross", "Cigna", "UnitedHealthcare", and "Medicare".
- Billing Amount: This column specifies the money billed for the patient's healthcare services during admission. This is expressed as a floating-point number.
- Room Number: This column specifies the room number where the patient was accommodated during admission.
- Admission Type: This column specifies the type of admission, which can be "Emergency", "Elective", or "Urgent", reflecting the circumstances of the admission.
- Discharge Date: This column specifies the date the patient was discharged from the healthcare facility based on the admission date and a random number of days within a realistic range.
- Medication: This column identifies a medication prescribed or administered to the patient during admission. Examples include "Aspirin", "Ibuprofen", "Penicillin", "Paracetamol", and "Lipitor".

- Test Results: This column describes the results of a medical test conducted during the patient's admission. Possible values include "Normal", "Abnormal", or "Inconclusive", indicating the outcome of the test.

4.1. Ethical Considerations in Using Synthetic Healthcare Data

While using synthetic data addresses some privacy concerns, it is important to consider the ethical implications:

- Representation and Bias: Synthetic data may not accurately represent the complexities and nuances of real patient data, potentially introducing or amplifying biases [66].
- Validity of Research: Results obtained from synthetic data may not be as reliable or generalizable as those from real data, raising ethical questions about the validity of any conclusions drawn [67].
- Transparency: It is crucial to be transparent about the use of synthetic data in any analysis or research to maintain scientific integrity [68].
- Data Quality: The quality of synthetic data depends on the algorithms and methods used to generate it. Ensuring high-quality synthetic data is an ethical responsibility to prevent misleading results [69].

4.2. Implementation of Data Privacy and Ethical Practices

In this case study, we will apply the following data privacy and ethical practices to the Healthcare Dataset:

1. Data Minimization: We will reduce the amount of personal information stored, focusing only on essential data for healthcare analysis.
2. Anonymization and Pseudonymization: We will apply techniques to remove or replace Personally Identifiable Information.
3. Data Encryption: Sensitive fields will be encrypted to protect against unauthorized access.
4. Data Hashing: We will implement hashing for data integrity verification.
5. Access Controls: We will establish role-based access controls to limit data access based on user roles and responsibilities.
6. Transparent Communication: We will document all data handling processes and make this information available to relevant stakeholders.

These practices align with the ethical considerations discussed earlier and comply with regulations like GDPR and HIPAA. In the next section, we will detail the implementation of these practices and analyze their impact on data privacy, security, and usability.

5. Results and Analysis

In this section, we apply and analyze the results obtained from implementing best practices in database management in our case study.

5.1. Data Minimization

To minimize the data in the dataset, we chose to reduce the information in the "Gender" and "Test Results" columns.

In the column referring to the patient's biological gender, we replaced the entries "Male" and "Female" with "M" and "F", respectively.

In the test results column, we replaced the categories "Normal", "Abnormal" and "Inconclusive" with "N", "A", and "I".

By minimizing the data in the columns referring to the patient's biological sex and test results, we obtained the results in Tables 7 and 8:

Table 7. Data before minimization.

Name	Gender	Test Results
Tiffany Ramirez	Female	Inconclusive
Ruben Burns	Male	Normal
Chad Byrd	Male	Normal
Antonio Frederick	Male	Abnormal

Table 8. Data after minimization.

Name	Gender	Test Results
Tiffany Ramirez	F	I
Ruben Burns	M	N
Chad Byrd	M	N
Antonio Frederick	M	A

The results obtained by implementing data minimization practices in the dataset allowed us to reduce the amount of sensitive information stored, limiting the risk of exposure and privacy violations. Additionally, we can avoid unnecessary overhead, storage, and processing time by reducing the amount of data collected.

Ethical Implications of Data Minimization

While data minimization effectively reduces privacy risks, it is important to consider its ethical implications:

- Data Utility vs. Privacy: By minimizing data, we may limit the potential for certain types of analysis or research. This presents an ethical dilemma between protecting individual privacy and potentially limiting beneficial uses of data [16].
- Potential Bias: Minimizing certain data fields (like gender) could potentially introduce or exacerbate bias in analyses if not carefully considered [70].
- Informed Consent: It is crucial to ensure that data subjects are aware of and consent to this type of data transformation, as it changes the nature of the data they initially provided [71].

5.2. Anonymization and Pseudonymization of Data

To anonymize and pseudonymize the data in the dataset, we chose to pseudonymize the information in the "Name" column and anonymize the information in the "Age", "DateOfAdmission", and "DischargeDate" columns.

In the column referring to the patient's name, we used the character replacement (masking) method to conceal the patient's last name, using unique values like "*" to avoid the direct identification of individuals. This method can be applied to either categorical or numerical attributes.

In the column referring to the age of the patients, we applied a generalization method that groups the ages into 10-year intervals/ranges. This anonymization practice reduces the granularity of the data, making it less likely to identify specific individuals. Following this, we adopted a similar generalization approach for patient admission and discharge dates, where only the month and year were kept while the day was removed.

By pseudonymizing the data in the column referring to the patient's name, we obtained the results in Tables 9 and 10:

Table 9. Name before pseudonymization.

Name	Gender
Tiffany Ramirez	F
Ruben Burns	M
Chad Byrd	M
Antonio Frederick	M

Table 10. Name after pseudonymization.

Name	Gender
Tiffany *******	F
Ruben *****	M
Chad ****	M
Antonio *********	M

This pseudonymization approach reduces the risk of re-identification while protecting patient privacy.

By anonymizing the data in the column referring to the patient's age using the generalization technique, we obtained the results in Tables 11 and 12:

Table 11. Age before generalization.

Name	Age
Tiffany Ramirez	81
Ruben Burns	35
Chad Byrd	61
Antonio Frederick	49

Table 12. Age after generalization

Name	Age
Tiffany Ramirez	80–89
Ruben Burns	30–39
Chad Byrd	60–69
Antonio Frederick	40–49

In this way, patient privacy is preserved without significantly compromising the usefulness of the data for statistical and trend analysis.

Next, by anonymizing the data in the columns referring to the patient's admission and discharge dates using the generalization technique, we obtained the outcomes in Tables 13 and 14:

Table 13. Dates before generalization.

Name	Date of Admission	Discharge Date
Tiffany Ramirez	17 November 2022	1 December 2022
Ruben Burns	1 June 2023	15 June 2023
Chad Byrd	9 January 2019	8 February 2019
Antonio Frederick	2 May 2020	3 May 2020

Table 14. Dates after generalization.

Name	Date of Admission	Discharge Date
Tiffany Ramirez	2022-11	2022-12
Ruben Burns	2023-06	2023-06
Chad Byrd	2019-01	2019-02
Antonio Frederick	2020-05	2020-05

This simplification of data helps protect patient privacy while enabling temporally relevant analysis.

Finally, by anonymizing the data in the columns referring to the patient's room number using the suppression technique, we obtained the results in Tables 15 and 16:

Table 15. Room number before suppression.

Name	Room Number	Admission Type
Tiffany Ramirez	146	Elective
Ruben Burns	404	Emergency
Chad Byrd	292	Emergency
Antonio Frederick	480	Urgent

Table 16. Room number after suppression.

Name	Admission Type
Tiffany Ramirez	Elective
Ruben Burns	Emergency
Chad Byrd	Emergency
Antonio Frederick	Urgent

The results obtained with the implementation of anonymization and pseudonymization of the data in the dataset made it possible to remove and modify information that could identify patients, guaranteeing the privacy and security of the data.

Data anonymization and pseudonymization provided additional protection, making it more difficult to identify individuals from the available data. By suppressing sensitive attributes and generalizing information such as age and dates, it was possible to preserve patient privacy without compromising data integrity for statistical and trend analysis. However, these anonymization and pseudonymization techniques have caused a reduction in data granularity, making detailed analysis or cross-referencing difficult.

Ethical Considerations in Anonymization and Pseudonymization

While these techniques enhance privacy, they also raise ethical concerns:

- Re-identification Risk: Despite our efforts, there is always a risk of re-identification, especially with advances in data analytics. This presents an ongoing ethical challenge [18].
- Data Utility vs. Privacy: The reduction in data granularity may limit certain types of analysis, potentially impacting the utility of the data for research or healthcare improvements [72].
- Informed Consent: It is crucial to consider whether the original consent given by data subjects covers these transformations, as they significantly alter the nature of the data [71].
- Transparency: We must be transparent about these processes to maintain trust with data subjects and other stakeholders [73].

5.3. Data Encryption

In pgAdmin4, we used the Advanced Encryption Standard (AES) algorithm with the help of the *pgcrypto* extension in PostgreSQL, ensuring that only authorized users with access to the private key could view or modify the information.

We executed the query "SELECT * FROM PatientRecord ORDER BY id" to analyze the database's performance and the response time in milliseconds. Before data encryption, we observed a response time of less than 1 s, as shown in Figure 5:

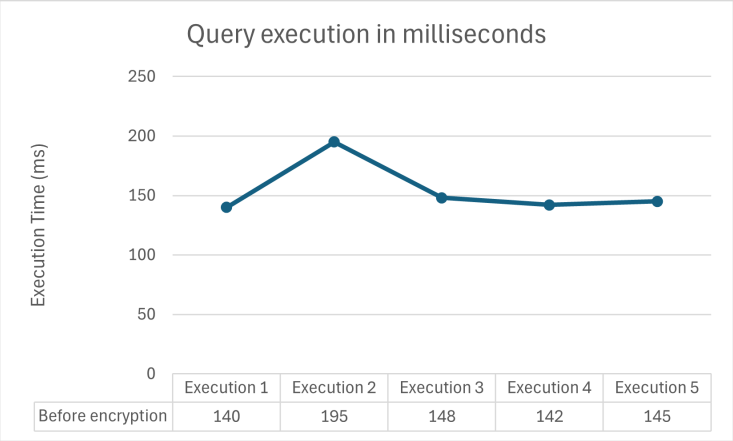

Figure 5. Query execution before encryption.

After encrypting the data in the columns referring to the patient's medical condition and medication, we can observe a response time of approximately 11 s, as illustrated in Figure 6:

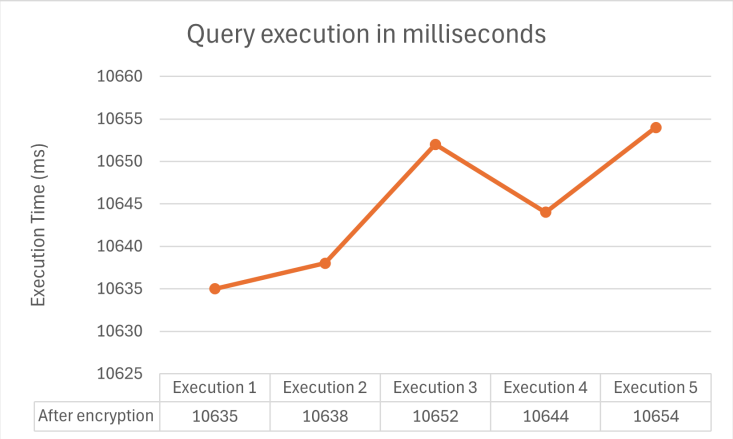

Figure 6. Query execution after encryption.

Data encryption resulted in a considerable increase in database response time, indicating a compromise between security and performance that must be carefully evaluated to ensure the appropriate balance between the two aspects. This suggests that encryption may harm system performance, although it provides greater security in protecting sensitive data, as seen in Figure 7.

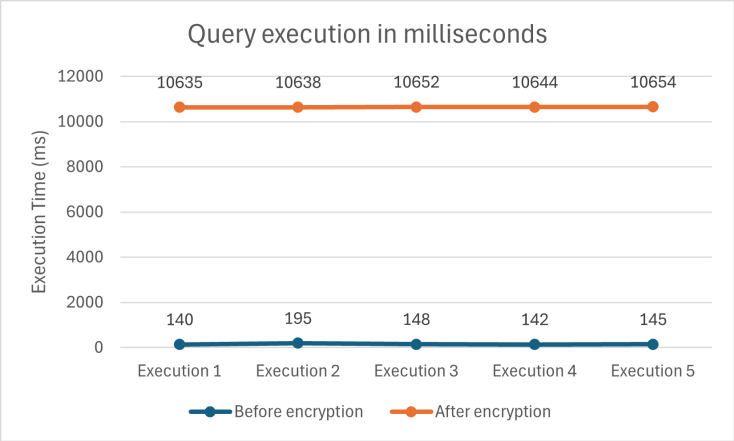

Figure 7. Query execution before and after encryption.

Data encryption, although essential for protecting privacy and data security, resulted in a loss in the performance of our database queries. This was because the additional time required to encrypt and decrypt information increased the system's response time, making it less efficient. While security is paramount, it is important to consider performance and find a balance between these opposing needs.

Ethical Implications of Data Encryption

The implementation of data encryption raises several ethical considerations:

- Security vs. Accessibility: While encryption enhances data security, it may limit data accessibility for legitimate uses, potentially impacting patient care or research [19].
- Performance Trade-offs: The significant increase in query execution time could affect the efficiency of healthcare operations, potentially impacting patient care. This presents an ethical dilemma between data security and operational efficiency [45].
- Key Management: The ethical responsibility of managing encryption keys is crucial, as loss of keys could result in permanent data loss [35].
- Transparency: It is important to be transparent about encryption practices to maintain trust with data subjects and comply with regulations [74].

5.4. Access Controls

To implement access control on data using pgAdmin4, we adopted an approach that involved assigning permissions to database users. This included granting specific access privileges, such as SELECT, INSERT, UPDATE, and DELETE, only to users who required these permissions to perform their functions.

Initially, the user (or role) "henrique" was created in PostgreSQL using the command "CREATE ROLE henrique LOGIN PASSWORD 'string' ". Then "henrique" was granted a role with "SELECT" and "INSERT" permissions for all tables in the public schema through the "GRANT" command. The "UPDATE" and "DELETE" permissions were revoked, using the "REVOKE" command to ensure more restricted data management. Furthermore, additional restrictions were applied to user "henrique", allowing him to authenticate and access the database using credentials. Some examples prevent him from having all administrative privileges (superuser), creating databases, creating new users (or roles), inheriting parental role privileges, and initiating streaming replication and backups.

User "afonso" received the "SELECT" permission for all tables in the public schema and the "INSERT", "UPDATE", and "DELETE" permissions were revoked. Likewise, additional restrictions were applied, very much similar to user "henrique".

Lastly, user "eduardo" was granted "SELECT", "INSERT", "UPDATE", and "DELETE" permissions for all tables in the public schema. Additional restrictions were applied to user "eduardo", allowing him to authenticate and access the database using credentials, preventing him from being a superuser, allowing him to create databases, create new

users or roles, inherit privileges from parental roles, and initiate streaming replication and backups.

These measures aim to guarantee precise and granular control of access to data, thus promoting the security and integrity of information stored in the health database against unauthorized access and improper manipulation.

Ethical Considerations in Access Control

Implementing access controls raises several ethical considerations:

- Balancing Access and Privacy: While restricting access enhances privacy, it may limit legitimate data use for research or patient care. This presents an ethical challenge in balancing data protection with data utility [52].
- Transparency and Accountability: It is crucial to maintain transparency about access control policies and ensure accountability for data access [75].
- Fairness in Access Distribution: The distribution of access rights must be fair and based on legitimate needs, avoiding any form of discrimination [53].
- Continuous Review: Access rights should be regularly reviewed to ensure they remain appropriate, reflecting the ethical responsibility to maintain data security over time [76].

In conclusion, implementing these data management practices has proven critical to ensuring compliance with privacy regulations, protecting the confidentiality of health information, and promoting user confidence in the ethical management and use of data. These measures contribute significantly to mitigating the risks of privacy violations and ensuring respect for patients' rights, even though they may minimally influence the speed and efficiency of the database.

5.5. Data Hashing

To implement data hashing, we used the SHA-256 algorithm provided by PostgreSQL's pgcrypto extension. We applied hashing to the patient's name and medical condition to verify data integrity.

Here is an example of how we implemented hashing:

```
UPDATE PatientRecord
SET NameHash = encode(digest(Name, 'sha256'), 'hex'),
    MedicalConditionHash = encode(digest(MedicalCondition, 'sha256'), 'hex');
```

This query creates hash values for the Name and MedicalCondition fields, storing them in the new columns NameHash and MedicalConditionHash, respectively.

To verify data integrity, we can compare the stored hash with a newly generated hash:

```
SELECT Name, NameHash,
       encode(digest(Name, 'sha256'), 'hex') AS ComputedHash,
       NameHash = encode(digest(Name, 'sha256'), 'hex') AS HashMatch
FROM PatientRecord;
```

This query computes a new hash for each Name and compares it with the stored hash, indicating whether they match.

Ethical Considerations in Data Hashing

While hashing enhances data integrity, it also raises ethical considerations:

- Privacy Implications: Although hashes are one-way functions, they could potentially be used for tracking or linking records across databases, raising privacy concerns [46].
- Transparency: It is important to be transparent about the use of hashing techniques to maintain trust with data subjects [77].
- Data Utility: Hashed data may limit certain types of data analysis or use, potentially impacting research or healthcare outcomes [51].

- Collision Risks: While rare with modern algorithms like SHA-256, hash collisions could potentially lead to data integrity issues, raising ethical concerns about data accuracy [50].

The implementation of data hashing provides an additional layer of security and data integrity verification, complementing our other data protection measures. However, it is crucial to balance these benefits against the ethical considerations and potential impacts on data utility.

6. Conclusions and Future Work

This study's implementation of data management practices proved essential to ensure that the health information in the "Healthcare Dataset" is adequately protected regarding privacy and security.

The implementation of data minimization, data anonymization and pseudonymization, data encryption, and access control has proven to be essential for effectively protecting the privacy and security of patients' health data.

Data minimization is essential to limit the amount of personal information collected and processed to the minimum necessary to achieve legitimate purposes, thus reducing the risks of exposure and misuse of data.

Anonymization, in turn, irreversibly eliminates the link between the data and the patient's identity, allowing the information to be used for statistical or research purposes without compromising privacy. Pseudonymization replaces direct identifiers with a code, maintaining the possibility of re-identification, which makes it suitable for situations where data needs to be traceable, such as monitoring the adverse effects of medications.

Encryption is essential in protecting patient data, providing additional protection against unauthorized access through reversible information encryption.

Access control establishes rules and restrictions on who can access, view, or modify data, ensuring that only authorized users can access and manipulate health data.

Data hashing provides an additional layer of security by ensuring data integrity and allowing for the verification of data authenticity. This is particularly important in healthcare settings where the accuracy of patient information is crucial.

While these security measures may introduce some performance overhead, particularly in the case of encryption, the privacy protection and regulatory compliance advantages are overwhelming. However, optimizing these implementations to minimize the performance impact and ensure an optimal balance between data security and system efficiency is essential.

The ethical implications of these practices are significant and multifaceted. They include considerations of data utility versus privacy, the potential for bias introduction, the need for informed consent, and the importance of transparency in data handling processes. These ethical considerations underscore the complexity of managing healthcare data and the need for a thoughtful, balanced approach that respects individual privacy while enabling the beneficial use of data for research and improved patient care.

Adopting these practices is crucial to building patient trust and promoting a safe, ethical, and trustworthy digital health ecosystem in digital information.

This study highlights the need for a comprehensive and ongoing approach to ensuring the privacy and security of health data in the digital era. It serves as a basis for future research and development in health information security.

Future Work

In future work, additional studies should determine the effectiveness of various data protection techniques by ensuring ongoing compliance with privacy regulations, investigating new anonymization and pseudonymization methods, and implementing monitoring and auditing systems.

Additionally, it would be interesting to consider ways to make the database structure more efficient, possibly looking to further minimize the negative impact of additional

security measures. These can be accomplished using more advanced data processing methods or optimizing database queries and operations.

Future research could also focus on the following:

- Exploring advanced encryption techniques that balance security and performance more effectively.
- Investigating the use of blockchain technology for enhancing data integrity and traceability in healthcare databases.
- Developing more sophisticated anonymization techniques that better preserve data utility while ensuring privacy.
- Studying the long-term impacts of these data protection measures on healthcare outcomes and research capabilities.
- Exploring the ethical implications of emerging technologies like artificial intelligence and machine learning in healthcare data management.
- Investigating ways to implement the "right to be forgotten" in healthcare databases while maintaining data integrity and complying with retention requirements.

Finally, it is imperative to stay updated with data privacy and security regulation changes, and monitor new technological developments and best practices in data management. These will ensure that the system remains adapted to the ever-evolving needs and requirements of the digital healthcare environment.

Author Contributions: Writing—original draft, E.P., J.R. and H.J.; Supervision, M.A. and P.M.; writing—review and editing, M.A., P.M., C.W., P.V. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by National Funds through the FCT—Foundation for Science and Technology, I.P., within the scope of the project Ref. UIDB/05583/2020. Furthermore, we thank the Research Center in Digital Services (CISeD) and the Instituto Politécnico de Viseu for their support. Maryam Abbasi thanks the national funding by FCT—Foundation for Science and Technology, I.P., through the institutional scientific employment program contract (CEECINST/00077/2021).

Data Availability Statement: Data is contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Quach, S.; Thaichon, P.; Martin, K.D.; Weaven, S.; Palmatier, R.W. Digital technologies: Tensions in privacy and data. *J. Acad. Mark. Sci.* **2022**, *50*, 1299–1323. [CrossRef]
2. Janic, M.; Wijbenga, J.P.; Veugen, T. Transparency enhancing tools (TETs): An overview. In Proceedings of the 2013 Third Workshop on Socio-Technical Aspects in Security and Trust, New Orleans, LA, USA, 29 June 2013; pp. 18–25.
3. IT Governance Privacy Team. *Eu General Data Protection Regulation (Gdpr)—An Implementation and Compliance Guide*; IT Governance Ltd.: Ely, Cambridgeshire, UK, 2020.
4. European Parliament and Council of the European Union. General Data Protection Regulation. *Off. J. Eur. Union* **2016**, *59*, L119/1–L119/88.
5. Act, A. Health Insurance Portability and Accountability Act. *Public Law* **2023**, *104*, 191.
6. Thapa, C.; Camtepe, S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Comput. Biol. Med.* **2021**, *129*, 104130. [CrossRef] [PubMed]
7. Kurteva, A.; Chhetri, T.R.; Pandit, H.J.; Fensel, A. Consent through the lens of semantics: State of the art survey and best practices. *Semant. Web* **2021**, *15*, 1–27.
8. Omotunde, H.; Ahmed, M. A Comprehensive Review of Security Measures in Database Systems: Assessing Authentication, Access Control, and Beyond. *Mesopotamian J. Cybersecur.* **2023**, *2023*, 115–133. [CrossRef] [PubMed]
9. Stair, R.M.; Reynolds, G.W. *Fundamentals of Information Systems*; Cengage Learning: Boston, MA, USA, 2018.
10. Hulkower, R.; Penn, M.; Schmit, C. Privacy and confidentiality of public health information. In *Public Health Informatics and Information Systems*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 147–166.
11. Patil, P. Healthcare Dataset, 2023. Available online: https://www.kaggle.com/datasets/prasad22/healthcare-dataset (accessed on 15 July 2024).
12. Ranganathan, R.; Kumar, G.S.; Angel, T.S. A Tool for Database Masking and Anonymization of PostgreSQL. In Proceedings of the 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), Chennai, India, 21–23 December 2023; pp. 1–6.

13. de Souza Rosa, A.; Lazarin, N.M. Uma análise de desempenho de funções de encriptação nativas de SGDBs Open Source. In Proceedings of the Anais Estendidos do XXIII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais, Juiz de Fora, Brazil, 18–21 September 2023; SBC: Porto Alegre, RS, Brasil, 2023; pp. 117–128.
14. Sharma, S.; Choudhary, N.; Jain, K. A Study on Models and Techniques of Anonymization in Data Publishing. *Int. J. Sci. Res. Sci. Eng. Technol. IJSRSET* **2019**, *6*, 84–90. [CrossRef]
15. Tahir, H.; Brezillon, P. Data Anonymization Process Challenges and Context. *Int. J. Database Manag. Syst. (IJDMS)* **2023**, *15*, 59–69. [CrossRef]
16. Galdon Clavell, G.; Martín Zamorano, M.; Castillo, C.; Smith, O.; Matic, A. Auditing Algorithms: On Lessons Learned and the Risks of Data Minimization. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, New York, NY, USA, 7–9 February 2020; pp. 265–271. [CrossRef]
17. Heuer, H.; Breiter, A. Student Success Prediction and the Trade-Off between Big Data and Data Minimization. In *DeLFI 2018—Die 16. E-Learning Fachtagung Informatik*; Gesellschaft für Informatik e.V.: Bonn, Germany, 2018; pp. 219–230.
18. Neubauer, T.; Heurix, J. A methodology for the pseudonymization of medical data. *Int. J. Med. Inform.* **2011**, *80*, 190–204. [CrossRef]
19. Nadeem, A.; Javed, M. A Performance Comparison of Data Encryption Algorithms. In Proceedings of the 2005 International Conference on Information and Communication Technologies, Karachi, Pakistan, 27–28 August 2005; pp. 84–89. [CrossRef]
20. Sarathy, R.; Robertson, C. Strategic and Ethical Considerations in Managing Digital Privacy. *J. Bus. Ethics* **2003**, *46*, 111–126. [CrossRef]
21. Shabani, M. The Data Governance Act and the EU's move towards facilitating data sharing. *Mol. Syst. Biol.* **2021**, *17*, e10229. [CrossRef]
22. Schreiber, K.; Pommerening, P.; Schoel, P. New Data Governance Act. A Practitioner's Guide. 2023. Available online: https://www.nomos-elibrary.de/10.5771/9783748937050/new-data-governance-act?page=1 (accessed on 20 July 2024).
23. Cauffman, C.; Goanta, C. A New Order: The Digital Services Act and Consumer Protection. *Eur. J. Risk Regul.* **2021**, *12*, 758–774. [CrossRef]
24. Veale, M.; Borgesius, F.Z. Demystifying the Draft EU Artificial Intelligence Act. *arXiv* **2021**, arXiv:2107.03721. [CrossRef]
25. Nair, S.R. A review on ethical concerns in big data management. *Int. J. Big Data Manag.* **2020**, *81*, 8. [CrossRef]
26. de Carvalho Ramos, A. The Right to Be Forgotten and the Indirect Control of Consumer Databases. In *Consumer Law and Socioeconomic Development: National and International Dimensions*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 371–381. [CrossRef]
27. Shah, W.F. Preserving Privacy and Security: A Comparative Study of Health Data Regulations—GDPR vs. HIPAA. *Int. J. Res. Appl. Sci. Eng. Technol.* **2023**, *11*, 55551. [CrossRef]
28. Mbonihankuye, S.; Nkunzimana, A.; Ndagijimana, A. Healthcare Data Security Technology: HIPAA Compliance. *Wirel. Commun. Mob. Comput.* **2019**, *2019*, 1927495:1–1927495:7. [CrossRef]
29. Chiou, L.; Tucker, C. *Search Engines and Data Retention: Implications for Privacy and Antitrust*; Working Paper 23815; National Bureau of Economic Research: Cambridge, MA, USA, 2017.
30. Blanchette, J.F.; Johnson, D.G. Data Retention and the Panoptic Society: The Social Benefits of Forgetfulness. *Inf. Soc.* **2002**, *18*, 33–45. [CrossRef]
31. Ausloos, J. Conditions of the Right to Erasure. In *The Right to Erasure in EU Data Protection Law*; Oxford University Press: Oxford, UK, 2020; pp. 196–274. [CrossRef]
32. Faisal, K. Applying the Purpose Limitation Principle in Smart-City Data-Processing Practices: A European Data Protection Law Perspective. *Commun. Law Policy* **2023**, *28*, 67–97. [CrossRef]
33. Shanmugam, D.; Shabanian, S.; Díaz, F.; Finck, M.; Biega, A.J. Learning to Limit Data Collection via Scaling Laws: Data Minimization Compliance in Practice. *arXiv* **2021**, arXiv:2107.08096.
34. Enzmann, M.; Selzer, A.; Spychalski, D. Practitioner's Corner Data Erasure under the GDPR—Steps towards Compliance. *Eur. Data Prot. Law Rev.* **2019**, *5*, 416–420. [CrossRef]
35. Naguib, A.; Fouad, K.M. Database Security: Current Challenges and Effective Protection Strategies. In Proceedings of the 2024 6th International Conference on Computing and Informatics (ICCI), New Cairo, Egypt, 6–7 March 2024; pp. 120–130.
36. Tyagi, A.K. *Privacy Preservation and Secured Data Storage in Cloud Computing*; IGI Global: Hershey, PA, USA, 2023.
37. Rouzrokh, P.; Khosravi, B.; Faghani, S.; Moassefi, M.; Garcia, D.V.V.; Singh, Y.; Zhang, K.; Conte, G.; Erickson, B. Mitigating Bias in Radiology Machine Learning: 1. Data Handling. *Radiol. Artif. Intell.* **2022**, *45*, e210290. [CrossRef] [PubMed]
38. Vovk, O.; Piho, G.; Ross, P. Methods and tools for healthcare data anonymization: A literature review. *Int. J. Gen. Syst.* **2023**, *52*, 326–342. [CrossRef]
39. Marques, J.F.; Bernardino, J. Analysis of Data Anonymization Techniques. In Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020)—Volume 2: KEOD, Virtual, 2–4 November 2020; pp. 235–241. [CrossRef]
40. Al-Zubaidie, M.; Zhang, Z.; Zhang, J. PAX: Using Pseudonymization and Anonymization to Protect Patients' Identities and Data in the Healthcare System. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1490. [CrossRef] [PubMed]
41. Xu, L.; Jiang, C.; Chen, Y.; Ren, Y.; Liu, K. Privacy or Utility in Data Collection? A Contract Theoretic Approach. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1256–1269. [CrossRef]

42. Pedrosa, M.; Zúquete, A.; Costa, C. A Pseudonymisation Protocol With Implicit and Explicit Consent Routes for Health Records in Federated Ledgers. *IEEE J. Biomed. Health Inform.* **2020**, *25*, 2172–2183. [CrossRef] [PubMed]
43. Domingo-Ferrer, J.; Muralidhar, K. New directions in anonymization: Permutation paradigm, verifiability by subjects and intruders, transparency to users. *arXiv* **2015**, arXiv:1501.04186. [CrossRef]
44. Boyd, C. Modern data encryption. *Electron. Commun. Eng. J.* **1993**, *5*, 271–278. [CrossRef]
45. Mustika, L. Implementasi Algoritma AES Untuk Pengamanan Login Dan Data Customer Pada E-Commerce Berbasis Web. *J. Ris. Komput.* **2020**, *7*, 148–155. [CrossRef]
46. Roshdy, R.; Fouad, M.; Aboul-Dahab, M. Design and Implementation a new Security Hash Algorithm based on MD5 and SHA-256. *Int. J. Eng. Sci. Emerg. Technol.* **2013**, *6*, 29–36.
47. MD5 LLC. MD5 Hashing, 2009. Online resource. Available online: https://www.md5online.org/ (accessed on 22 July 2024).
48. Sadeghi-Nasab, A.; Rafe, V. A comprehensive review of the security flaws of hashing algorithms. *J. Comput. Virol. Hacking Tech.* **2023**, *19*, 287–302. [CrossRef]
49. Turner, S.; Chen, L. Updated Security Considerations for the MD5 Message-Digest and the HMAC-MD5 Algorithms. 2011. Available online: https://www.rfc-editor.org/rfc/rfc6151.html (accessed on 25 July 2024).
50. Prasanna, S.R.; Premananda, B. Performance Analysis of MD5 and SHA-256 Algorithms to Maintain Data Integrity. In Proceedings of the 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India, 27–28 August 2021; pp. 246–250. [CrossRef]
51. Rabtsani, M.R.; Triayudi, A.; Soepriyono, G. Combination of AES (Advanced Encryption Standard) and SHA256 Algorithms for Data Security in Bill Payment Applications. *SAGA J. Technol. Inf. Syst.* **2024**, *2*, 175–189. [CrossRef]
52. Ryan, J. Information security tools and practices: What works? *IEEE Trans. Comput.* **2004**, *53*, 1060–1063. [CrossRef]
53. Tolone, W.; Ahn, G.J.; Pai, T.; Hong, S.P. Access control in collaborative systems. *ACM Comput. Surv.* **2005**, *37*, 29–41. [CrossRef]
54. Lane, J.; Schur, C. Balancing access to health data and privacy: A review of the issues and approaches for the future. *Health Serv. Res.* **2010**, *45*, 1456–1467. [CrossRef]
55. Schnackenberg, A.; Tomlinson, E.C. Organizational Transparency. *J. Manag.* **2016**, *42*, 1784–1810. [CrossRef]
56. Hacker, P. Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Mark. Law Rev.* **2018**, *55*, 1143–1185. [CrossRef]
57. Sundareswaran, S.; Squicciarini, A.; Lin, D. Ensuring Distributed Accountability for Data Sharing in the Cloud. *IEEE Trans. Dependable Secur. Comput.* **2012**, *9*, 556–568. [CrossRef]
58. International Agency for Research on Cancer. Effective and transparent communication with target populations and other stakeholders. In *Best Practices in Cervical Screening Programmes: Audit of Cancers, Legal and Ethical Frameworks, Communication, and Workforce Competencies*; IARC Working Group Reports, No. 11; International Agency for Research on Cancer: Lyon, France 2023; Chapter 3.
59. Holt, V.; Ramage, M.; Kear, K.; Heap, N. The usage of best practices and procedures in the database community. *Inf. Syst.* **2015**, *49*, 163–181. [CrossRef]
60. Semantha, F.H.; Azam, S.; Shanmugam, B.; Yeo, K.C. PbDinEHR: A Novel Privacy by Design Developed Framework Using Distributed Data Storage and Sharing for Secure and Scalable Electronic Health Records Management. *J. Sens. Actuator Netw.* **2023**, *12*, 36. [CrossRef]
61. Rantung, V.P.; Kainde, Q.C. Database design for agile stakeholder communication. In Proceedings of the 2015 1st International Conference on Wireless and Telematics (ICWT), Manado, Indonesia, 17–18 November 2015; pp. 1–5. [CrossRef]
62. Auger, G.A. Trust Me, Trust Me Not: An Experimental Analysis of the Effect of Transparency on Organizations. *J. Public Relat. Res.* **2014**, *26*, 325–343. [CrossRef]
63. Grady, C. Enduring and emerging challenges of informed consent. *N. Engl. J. Med.* **2015**, *372*, 2172. [CrossRef] [PubMed]
64. Ortega-Rodríguez, C.; Licerán-Gutiérrez, A.; Moreno-Albarracín, A.L. Transparency as a Key Element in Accountability in Non-Profit Organizations: A Systematic Literature Review. *Sustainability* **2020**, *12*, 5834. [CrossRef]
65. Psoinos, A.; Kern, T.; Smithson, S. An exploratory study of information systems in support of employee empowerment. *J. Inf. Technol.* **2000**, *15*, 211–230. [CrossRef]
66. Micheletti, N.; Marchesi, R.; Kuo, N.I.H.; Barbieri, S.; Jurman, G.; Osmani, V. Generative AI Mitigates Representation Bias Using Synthetic Health Data. *medRxiv* **2023**. [CrossRef]
67. Azizi, Z.; Zheng, C.; Mosquera, L.; Pilote, L.; Emam, K.E. Can synthetic data be a proxy for real clinical trial data? A validation study. *BMJ Open* **2021**, *11*, e043497. [CrossRef] [PubMed]
68. Iqbal, S.; Wallach, J.; Khoury, M.; Schully, S.; Ioannidis, J. Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biol.* **2016**, *14*, e1002333. [CrossRef] [PubMed]
69. Chauhan, P.; Bongo, L.A.; Pedersen, E. Ethical Challenges of Using Synthetic Data. *Proc. Aaai Symp. Ser.* **2023**. [CrossRef]
70. Bhardwaj, R.; Majumder, N.; Poria, S. Investigating Gender Bias in BERT. *Cogn. Comput.* **2020**, *13*, 1008–1018. [CrossRef]
71. Kaye, J.; Whitley, E.; Lund, D.; Morrison, M.; Teare, H.J.A.; Melham, K. Dynamic consent: A patient interface for twenty-first century research networks. *Eur. J. Hum. Genet.* **2014**, *23*, 141–146. [CrossRef]
72. Yale, A.; Dash, S.; Dutta, R.; Guyon, I.M.; Pavao, A.; Bennett, K.P. Generation and evaluation of privacy preserving synthetic health data. *Neurocomputing* **2020**, *416*, 244–255. [CrossRef]

73. Esteves, B.; Asgarinia, H.; Penedo, A.C.; Mutiro, B.; Lewis, D. Fostering trust with transparency in the data economy era: An integrated ethical, legal, and knowledge engineering approach. In Proceedings of the 1st International Workshop on Data Economy, Rome, Italy, 9 December 2022. [CrossRef]
74. Zhan, W.D.; Jin, B.; Xu, H.; Dong, C. Data Security Management Based on Transparent Encryption Policy. In Proceedings of the 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2–3 December 2022; pp. 1–4. [CrossRef]
75. Kunduru, A.R. Industry best practices on implementing oracle cloud ERP security. *Int. J. Comput. Trends Technol.* **2023**, *71*, 1–8. [CrossRef]
76. Kaigai. SEPostgreSQL Introduction. Online resource. Available online: https://github.com/kaigai/sepostgresql (accessed on 25 July 2024).
77. Bertino, E.; Kundu, A.; Sura, Z. Data Transparency with Blockchain and AI Ethics. *J. Data Inf. Qual. (JDIQ)* **2019**, *11*, 1–8. [CrossRef]