Utilizing blockchain technologies for ensuring the confidentiality and security of personal data in compliance with GDPR

Valeriia Balatska[1,†], Vasyl Poberezhnyk[1,†] and Ivan Opirskyy[1,*,†]

[1] Lviv Polytechnic National University, 12 Stepana Bandery str., 79000 Lviv, Ukraine

Abstract

The growth of the requirements for personal data protection according to the General Data Protection Regulation (GDPR) leads to the increasing cost of developing innovative solutions to ensure the privacy and security of information. The conception of the service for collecting and storing personal data that uses blockchain technologies, smart contracts, NFTs (non-fungible tokens), and the decentralized file system IPFS (InterPlanetary File System) is proposed in this paper. The main functional blocks of the service such as user registration and authentication, data collection and processing, consent management, and access tracking are considered. Blockchain usage allows the immutability and transparency of user accounts and consents, while smart contracts automate the process of managing access rights. The use of IPFS for decentralized file storage increases security and ensures data integrity. Data processing techniques that include masking, pseudo-anonymization, shuffling, and perturbation, would significantly reduce the risk of disclosure of sensitive information. The proposed concept demonstrates the possibility of providing the effective protection of personal data that would be compliant with the GDPR while using modern technologies. The purpose of this work is the development of a concept of a service for the collection and storage of personal data that meets the requirements of the General Data Protection Regulation (GDPR) while using innovative technologies such as blockchain, smart contracts, NFT, and the IPFS.

Keywords

personal data protection, GDPR, blockchain, smart contracts, decentralized storage systems, IPFS, authorization, identification, access control, decentralization

1. Introduction

Today the processing of personal data in the digital world is a crucial part of many organizations' operations. With the development of technologies and the growth of the volume of data, the need to ensure reliable protection of confidential information also increases. The introduction of the General Data Protection Regulation (GDPR) in 2018 was an important step towards establishing high standards of personal data protection in the European Union and beyond. GDPR obliges organizations to implement appropriate measures to protect user data, which includes not only technical but also organizational aspects [1].

Despite the existence of many methods and tools that provide data security, new challenges that are related to cybercrime, data leaks, and privacy breaches are continuously emerging. This is the reason why the search for innovative approaches and technologies for effective information protection is crucial. In this scope, the usage of blockchain technology, which ensures the immutability and transparency of records, is promising, and the mentioned properties are critical for the protection of confidential data. Blockchain technologies can provide significant advantages in the scope of data privacy and security. Decentralized structure and cryptographic mechanisms provide the opportunity to avoid a single vulnerable point and ensure high resistance to unauthorized access while using blockchain technologies [2]. Moreover, blockchain helps to increase the level of trust between parties by ensuring the immutability of records and the possibility of public verification.

Smart contracts that operate based on the blockchain allow the automatization of access rights management and user consent, which will significantly reduce the risk of human errors and increase the efficiency of data management. The usage of the IPFS (InterPlanetary File System)—the decentralized file storage, ensures a high level of data security, integrity, and availability. The concept of the service is designed for collecting and storing personal data by GDPR requirements using modern technologies such as blockchain, smart contracts, NFT, and the decentralized IPFS file system. The proposed approach includes various data processing methods such as masking, pseudo-anonymization, shuffling, and perturbation, which can significantly reduce the risk of revealing confidential information.

CSDP-2024: Cyber Security and Data Protection, June 30, 2024, Lviv, Ukraine
*Corresponding author.
†These authors contributed equally.
valeriia.s.balatska@lpnu.ua (V. Balatska); vasyl.poberezhnyk@gmail.com (V. Poberezhnyk); iopirsky@gmail.com (I. Opirskyy)
0000-0002-6262-6792 (V. Balatska); 0000-0002-7523-2557 (V. Poberezhnyk); 0000-0002-8461-8996 (I. Opirskyy)

The implementation of such a system contributes to increasing the level of personal data security and ensuring compliance with GDPR requirements, which is an urgent task for many organizations in the modern world. This work explores the possibilities of using innovative technologies to create secure and transparent information systems that meet modern regulatory requirements and challenges related to personal data protection.

Problem formulation. In today's digital world, where the personal data of users is becoming increasingly valuable, the protection of this data becomes a critical task for organizations. The requirements of the General Data Protection Regulation (GDPR) impose an obligation on businesses to ensure compliance with the management and protection of personal data, and violations of these requirements can lead to serious fines and loss of consumer trust.

However, existing methods of data protection are often insufficiently effective in ensuring full confidentiality and security. Traditional centralized data storage systems are vulnerable to cyber-attacks, data leaks, and unauthorized access. For example, in 2023 there was a major data breach that affected millions of users, highlighting the ineffectiveness of traditional protection methods.

This problem becomes even more urgent due to the rapid growth of the amount of personal data collected and processed by organizations. Every day, millions of new users register on online platforms, leaving their data, which makes the need for innovative solutions to protect this data increasingly acute.

The need for innovative solutions to protect personal data is becoming more and more urgent. Modern technologies such as blockchain, smart contracts, NFTs, and decentralized storage systems have the potential to radically change approaches to data security and privacy. The use of blockchain, for example, can ensure the immutability and transparency of records, while smart contracts can automate data access control and enforce privacy policies.

Thus, the key challenge is to develop and implement effective and innovative personal data collection and storage systems that meet GDPR requirements and provide the highest level of privacy and security. This paper partially addresses this issue by looking at how innovative technologies can help protect personal data, ensure GDPR compliance, and increase security in the digital environment.

Recent research and publications analysis. Recent research and publications in the field of personal data protection indicate a growing interest in the use of blockchain technologies to ensure the confidentiality and security of information [3]. Scientists and specialists are actively exploring the possibilities of blockchain to create decentralized data storage systems that can provide a high degree of protection against unauthorized access and data leaks. One of the main advantages of the blockchain is its immutability and transparency, which allows you to accurately track all actions with data and ensure their integrity.

Research suggests that blockchain can ensure data immutability, making it inaccessible to unauthorized editing. This is especially important for tracking all data actions and ensuring data integrity. Also, the blockchain allows the use of smart contracts to automate data access management processes. Smart contracts can play the role of automated agents that control the provision of access to personal data by established rules and user consent [4]. This increases the accuracy and security of data processing, minimizing the human factor and possible errors.

An important topic of research is the use of decentralized file systems, such as IPFS (InterPlanetary File System), to store large volumes of personal data. IPFS allows you to store data in a distributed environment, which increases their availability and protection against malicious attacks. Hashing data in IPFS ensures its integrity, which is an important aspect of GDPR compliance [5].

Immutability and transparency are key characteristics of blockchain technologies that make them particularly valuable for protecting personal data and other sensitive information assets. One of the main reasons for the immutability of data in a blockchain is its consensus mechanism. Blockchain is based on a distributed system where each node (computer) of the network has a copy of all transactions that have ever been made. These nodes use a consensus algorithm (such as Proof of Work, Proof of Stake, etc.) to agree on which blocks (groups of transactions) are added to the chain of blocks (blockchain) [6]. Only once a block is added to the blockchain is it almost impossible to change or remove it without the consensus of all other nodes, making the blockchain immutable.

Each block in the blockchain has a unique identifier, which is known as a hash. This value is generated from the block data using cryptographic hash functions such as SHA-256. If even one thing changes in a block (even one character), it will cause its hash value to change. Thus, even a small change in information will lead to a change in the hash, which will be visible to all network participants.

Also important is the immutability of blocks, once a block is added to the blockchain and becomes part of the overall sequence of blocks, it stays there forever. This creates immutability, or immutability, of data. This is especially important for tracking and historical verification of transactions in public blockchains [7].

Most blockchains, in particular public ones, have open access to all data stored in the network. This means that any member of the network can verify and view any transaction or account balance using a public key (address) in the network. Blockchain transparency is based on its decentralized nature. No centralized authority or authority controls the entire network. Instead, every node in the network has the same right to access and control the data, which ensures discovery verification and transaction transparency. Also, blockchain creates a reliable basis for open and transparent systems. In this regard, the values and principles underlying blockchain technologies make them an important tool for solving problems in the fields of finance, logistics, medicine, voting, and, of course, personal data protection [8]. These aspects of immutability and transparency give blockchains the power to protect data and increase trust in the digital systems that use them to operate. In addition to blockchain technologies, privacy-friendly data processing methods such as masking, pseudo-anonymization, shuffling, and perturbation are being

actively researched. These methods make it possible to ensure a high degree of confidentiality of personal data while maintaining their usefulness for analysis and other purposes. The use of such methods is key to complying with GDPR data protection requirements. In general, recent research and publications indicate the significant potential of blockchain technologies for creating secure and efficient systems for storing and processing personal data [9]. They highlight the importance of innovative approaches to data protection that can ensure a high level of privacy and security in the face of growing cybersecurity threats. The integration of blockchain technologies, smart contracts, IPFS, and advanced data processing methods is a promising direction to achieve GDPR compliance and increase the level of security in the digital environment.

The purpose of the paper is to develop the concept of an innovative service for collecting and storing personal data that meets the requirements of the General Data Protection Regulation (GDPR). The main objectives of the paper are:

- Analysis of existing personal data protection technologies, in particular, research modern methods and technologies of personal data protection used in the context of GDPR requirements, analysis of advantages and disadvantages of existing solutions, including centralized and decentralized data storage systems.
- Development of a service architecture for data collection and storage, to develop a service architecture that includes components for registration, authentication, collection, processing, and storage of personal data. Integrate blockchain to ensure immutability and transparency of records of user data and consent.
- Research on the use of smart contracts for managing access rights and consents. Develop models of smart contracts to automate the management of data processing consents. Explore the possibilities of using smart contracts to control access rights to personal data.
- Explore the possibilities and limitations of using IPFS for secure storage of large volumes of personal data. Develop mechanisms for IPFS integration with blockchain and smart contracts.
- Data processing with confidentiality in mind. Develop and implement data processing techniques such as masking, pseudo-anonymization, scrambling, and perturbation to ensure the privacy of personal data. Assess the effectiveness of these methods in the context of maintaining privacy and GDPR compliance.
- Assess the proposed service for compliance with GDPR requirements. Determine how the proposed technologies can help organizations meet regulatory requirements for the protection of personal data.
- Identify potential risks and challenges associated with the implementation of the proposed technologies. Develop recommendations for overcoming identified problems and reducing risks.

These tasks are aimed at creating an innovative service that will ensure a high level of personal data protection by GDPR requirements, using modern technologies to increase security and transparency.

2. Protection technologies and personal data processing architecture

2.1. Analysis of existing personal data protection technologies

The protection of personal data is critically important in today's digital world, especially with the requirements of the General Data Protection Regulation (GDPR) in the European Union. The GDPR sets high standards for the processing of personal data, in particular regarding their storage, transfer, and access [10]. In this regard, organizations must implement modern methods and technologies to ensure the confidentiality, integrity, and availability of personal data.

The main methods of protecting personal data are encryption, anonymization, pseudo-anonymization, access control, use of smart contracts, monitoring and auditing, decentralized data storage, masking, shuffling, and data perturbation [11]. Each of these methods has its advantages and disadvantages, which should be considered when choosing the appropriate solution for a specific organization.

Below is a table that contains a description of the main methods of personal data protection, their advantages and disadvantages.

From the analysis of the table, which contains a description of the main methods of personal data protection, their advantages and disadvantages, several important conclusions can be drawn.

First, each of the methods has its unique features and is applied depending on the specific needs of the organization and the nature of the data [12]. Encryption provides a high level of security and is indispensable for protecting data both at rest and in transit. However, high key management requirements can be difficult to implement.

Anonymization and pseudo-anonymization effectively ensure data confidentiality but may reduce their usefulness for further analysis [13]. Access control and smart contracts allow detailed configuration of access rights and automation of user consent management, but their implementation can be complex and require significant resources [14].

Monitoring and auditing are essential for data transparency and rapid incident response, although they require significant implementation and maintenance costs. Decentralized storage systems such as blockchain and IPFS offer high resistance to attacks and no single point of failure but are also characterized by high implementation costs and management complexity.

Data masking, shuffling, and perturbation provide an additional layer of privacy protection but may affect the accuracy and usefulness of the data.

Table 1
Methods of protecting personal data, their advantages and disadvantages

Method of protection	Description	Advantages	Disadvantages
Encryption	Converting data into an obscure format that is only available with a key	High level of security, protection against unauthorized access	High requirements for key management, significant processing costs
Anonymization	Complete removal of identifying information from data	High level of privacy, GDPR compliance	Decreased usefulness of data for analysis, the complexity of implementation
Pseudo anonymization	Replaces identifying data with artificial ones, making it difficult to identify a person	High level of confidentiality, preservation of partial usefulness of data	The possibility of restoring the identifier if additional data is available
Access control	Regulation of data access rights through multi-factor authentication and role management	Enhanced security, fine-tuning of access rights	Complexity of administration, need for constant monitoring
Smart contracts	Automation of user consent management	Automation, relevance of information, transparency	Complexity of implementation, need for a reliable infrastructure
Monitoring and auditing	Data access tracking and recording of all actions	Transparency of actions with data, quick response to incidents	High cost of implementation and maintenance, a large volume of data for analysis
Decentralized storage (IPFS, blockchain)	Distribution of data across many network nodes	High resistance to attacks, no single point of failure	Management complexity, high implementation cost, scalability issues
Data masking	Hiding the true values of the data while preserving its structure	Increased confidentiality, preserving the usefulness of data	It can reduce the accuracy of analysis, the complexity of implementation
Data shuffling	Changing the order of elements in a data array	An additional level of privacy protection	Can complicate data analysis, and requires additional resources

Summing up, the choice of a specific method of protecting personal data depends on the balance between the level of necessary security and the convenience of data use. Organizations should carefully assess their needs and opportunities for implementing certain technologies to ensure the maximum level of personal data protection and compliance with GDPR requirements.

2.2. Development of service architecture for collecting and storing personal data

To ensure compliance with the requirements of the General Data Protection Regulation (GDPR), a service architecture is proposed for the collection and storage of personal data using modern technologies, including blockchain, smart contracts, and the decentralized IPFS file system [15]. The architecture includes the following main components: registration, authentication, data collection, data processing, data storage, and consent management.

Service architecture:

- The registration and authentication component includes the collection of basic data to create a user profile, as well as the use of Multi-Factor Authentication (MFA) to ensure access security.
- Data collection component. The data collection interface includes forms and APIs for user input. The next step is to check the entered data for correctness and compliance with the requirements, i.e. data validation.
- Data processing component. This step involves anonymization and pseudo-anonymization, removal or replacement of identifying data to protect privacy, and data masking and perturbation.
- Data storage component. Blockchain integration, IPFS (InterPlanetary File System), and traditional databases are used to ensure immutability and transparency in the storage and processing of personal data, each of which has its advantages and applications in the service architecture.
- Consent management component. Smart contracts automate the process of obtaining and storing user consent, and an audit log records all data actions to ensure transparency and GDPR compliance.

Blockchain integration ensures the immutability of records by storing transactional data, including user consent to the processing of their data. This ensures that no record can be modified or deleted [16]. In addition, every entry in the blockchain is verifiable, which promotes a high level of trust from users and regulatory authorities. The use of smart contracts allows you to automate the management of consents and the implementation of privacy policies, which increases efficiency and reduces the risk of errors since they ensure the fulfillment of defined conditions without the need for human intervention.

The architectural diagram illustrates the components of the service for collecting and storing personal data. In this architecture:

1. Registration and authentication are provided through the user interface and the authentication server.
2. Data collection takes place using web forms and APIs, which are checked by a validation system.

3. Data processing includes anonymization, pseudo-anonymization, and other processing methods to ensure confidentiality.
4. Data is stored on a blockchain for immutability, on IPFS nodes for file integrity, and in a relational database for metadata.
5. Consent management is automated using smart contracts, and an audit log records all data actions.

The proposed architecture ensures a high level of security and privacy of personal data, considering GDPR requirements, thanks to the integration of blockchain, IPFS, and smart contracts.

The developed architecture of the personal data protection system ensures a high level of security and confidentiality of user information, combining innovative technologies with advanced data processing methods. The system includes several key components, each of which performs specific functions to ensure reliable data protection.

The registration and authentication component plays an important role in securing access to the system. It implements the creation and management of user profiles using MFA [17], which includes, for example, a combination of passwords and biometrics. This allows for a significant increase in the level of security, reducing the risks of unauthorized access. Databases store user information, and security protocols such as OAuth 2.0 and OpenID Connect provide secure authentication.

The data collection component includes interfaces that allow users to enter data through forms and APIs. Entered data undergoes thorough validation for correctness and compliance with requirements, which reduces the risk of errors and ensures high data quality. This component integrates with the front end (HTML, CSS, JavaScript) and RESTful APIs that allow efficient interaction with the system.

The data processing component uses anonymization and pseudo-anonymization techniques to remove or replace identifying data that protects user privacy. This also includes data masking and perturbation techniques that preserve the usefulness of the data for analysis while minimizing the risks of leaking sensitive information. Libraries such as Pandas (Python) and DataFrame (JavaScript) are used for data processing.

The data storage component provides the use of blockchain, IPFS (InterPlanetary File System), and traditional databases to store personal data. Blockchain ensures the immutability and transparency of records, allowing accurate tracking of all data activities, which is critical for GDPR compliance. IPFS allows you to store data in a distributed environment, increasing its availability and protection against malicious attacks. Traditional databases such as SQL and NoSQL are used to store structured data with high access speed.

The consent management component automates the process of obtaining and storing user consent using smart contracts. Smart contracts written in languages such as Solidity for Ethereum [18] act as automated agents that control access to data according to established rules and user consent. An audit log, which is kept recording all data actions, ensures transparency and compliance with GDPR requirements.

Compared to traditional centralized data storage systems, the developed architecture has several key advantages. It provides a higher level of security thanks to the use of blockchain and data anonymization methods, which reduces the risks of unauthorized access and data leaks. Transparency of data processing processes is achieved through blockchain records and an audit log, facilitating GDPR compliance. The use of smart contracts automates consent management and data processing processes, increasing efficiency and reducing the risk of errors.

Thus, the developed architecture presented in Fig. 1 is more effective and safer for protecting personal data, meeting GDPR requirements, and ensuring a high level of privacy in the face of growing cyber security threats.

The data storage system is made using a combination of blockchain, NFT, and IPFS technologies and plays the role of the core of the entire system, which is responsible for the secure storage of user data and ensuring access to this data. One of the key points is that this system must use a hybrid type of blockchain, as it is this type that allows for the possibility of delimiting access to blockchain records, as they contain user identifiers, therefore, to meet GDPR requirements, access to such data must be limited.

IPFS is used to store user data associated with an ID in a location outside the blockchain network to ensure the blockchain network's speed is maintained and its size minimized, as any excess data on the blockchain will increase its size and decrease its speed. Another, no less important criterion for the need to use the IPFS system is the need to ensure the possibility of deleting user data from the system if necessary. Since one of the properties of the blockchain is the impossibility of deleting data from it, the only possible option to ensure the possibility of deleting such data is to store them outside the blockchain network. Therefore, saving personal information in IPFS allows you to ensure that information is saved in a protected form from changes, as well as to ensure the possibility of deleting data from the system if necessary.

The last element that connects these technologies into one data accounting system is NFT (Non-Fungible Tokens) [19]. This innovative technology opens new perspectives for the identification and management of unique digital assets in the blockchain. In the context of personal data storage and management, NFT allows the creation of unique identifiers that can be associated with specific data sets in the decentralized file system IPFS (InterPlanetary File System). This ensures not only the tracking of data integrity and its immutability but also provides the ability to manage access to this data using smart contracts.

The use of NFT in the context of personal data protection allows you to enter additional parameters and metadata for each digital asset. For example, this may involve setting access rights at different levels for different users or groups of users. In addition, NFT can reflect data ownership rights, information processing conditions, and data access and use rules, which allows the creation of a more detailed and flexible data management system.

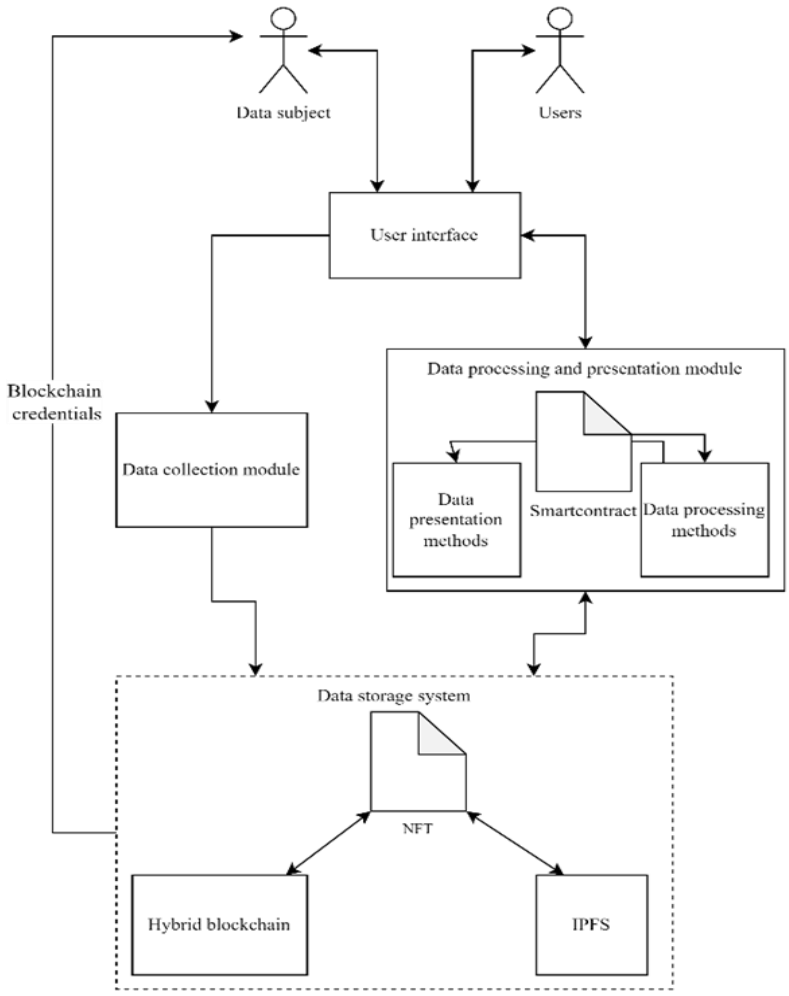

Figure 1: Architecture diagram

This approach significantly increases the level of security of personal data, because each element of information can be cryptographically protected and has clearly defined access conditions. Using NFT allows you to make the data management process more automated, efficient, and compliant with modern privacy protection requirements and regulatory standards, particularly GDPR.

The next element that should ensure the protection of personal data in this system is the data processing and presentation module. Its functioning is based on the use of smart contract technology. As you know, smart contracts are self-executing code in the blockchain network that executes when certain criteria are met. Accordingly, their use in this module allows you to avoid the human factor and manage access to information according to predetermined procedures, to ensure trust in such a mechanism, the code of such a smart contract must be publicly available, which again indicates the need to use a hybrid blockchain network. Because it gives the possibility to control the access to the information that is processed, for example, it will allow hiding user IDs, but it will allow access to the code of the smart contract, which will allow users to be assured of its reliability.

This smart contract should manage access to user data and ensure that information is provided to users in the form to which they have the right to access. Also, if necessary, grant the right to change information if the system user has the appropriate privileges.

In this approach, information is processed according to predefined principles in smart contracts. The use of this method allows to ensure trust in the system through the transparency of information processing mechanisms and collegial control over them.

Also, the presence of personal data in the information circulating in the network indicates that the use of an open-type blockchain is impractical, therefore the basis of such a system will be a hybrid-type blockchain network that allows delimiting access to information that circulates in it, which will allow delimiting identifiers users and other personal information, from other information that may be in the public domain, such as the previously mentioned smart contracts. Accordingly, the use of a hybrid blockchain will allow for maintaining trust in information processing mechanisms, leaving them in public access, and at the same time ensure the concealment of network user data while maintaining access to them only for predefined categories of network users.

Similar approaches are described in [20-21].

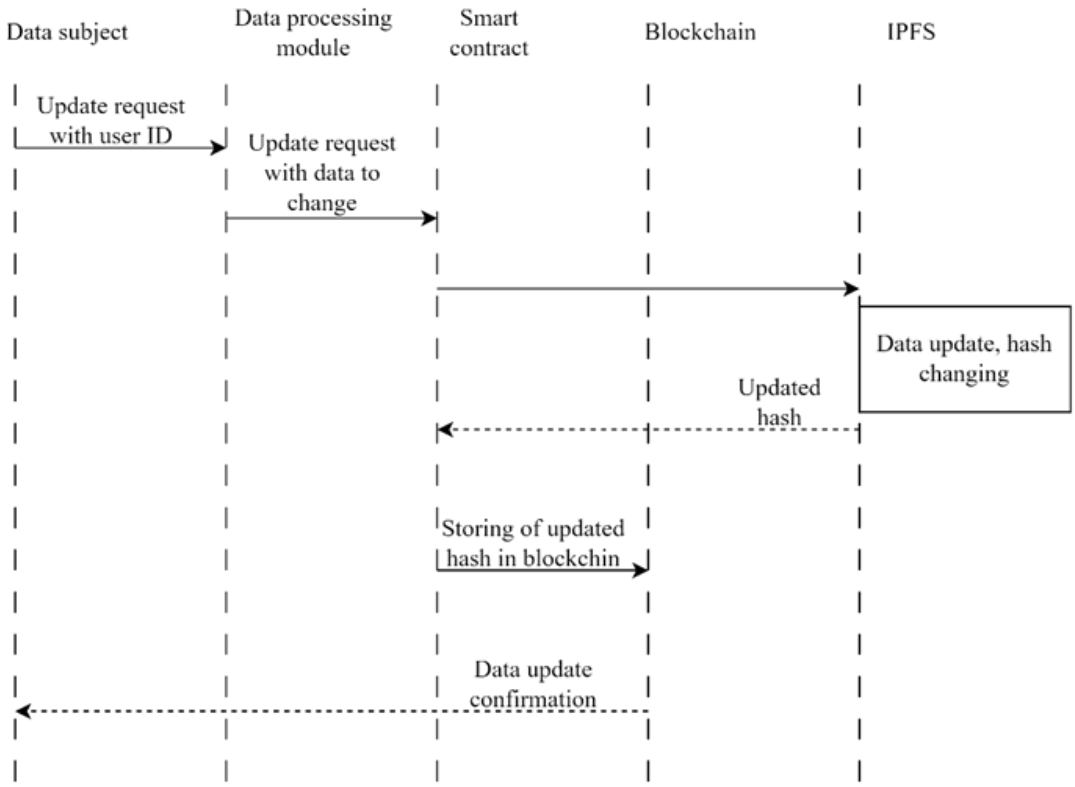

Figure 2: An example of updating user data in the system

In Fig. 2 the flow of updating or creating new data entry is depicted.

The algorithm in this case would go through these steps:

1. The data subject wants to update existing data.
2. The data subject goes to the processing module and fills in data to update and user ID.
3. The processing module forwards the request to the smart contract.
4. Smart contract checks data subject access rights and if successful forwards request to IPFS.
5. IPFS updates necessary data updates the hash of data entry, and returns it to the smart contract.
6. Smart contract updates the entry in NFT that exists in the blockchain.
7. Users receive the updated NFT from the blockchain.

In this scenario, the smart contract is responsible for managing the access rights to data modification that is stored in the IPFS. The usage of NFTs is required to provide opportunities for data modification, as usage of pure blockchain is impossible due to data immutability.

On the other hand, the request for data reading will be quite close to the updated one. However, it will have a different flow after the read request is obtained by a smart contract. The Fig. 3 depicts the flow.

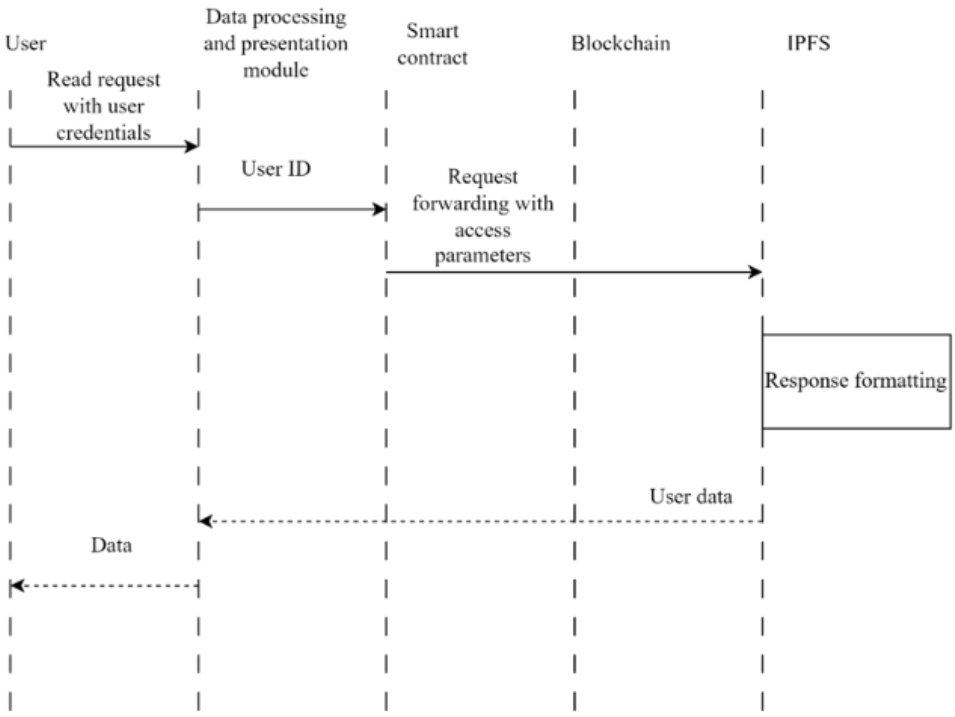

Figure 3: An example of passing a request to read data

The request to read data will go through the following stages:

1. The system user forms a request for access to information and submits it together with his identifier to the data processing and presentation module.
2. The data presentation module provides the user ID to the smart contract, which checks the user's access rights forms an access request according to the user's status, and passes it to IPFS.
3. IPFS forms a sample of data according to the given parameters and passes the sample back to the data processing and presentation module.
4. The sample is transferred to the user.

In this approach, access control is provided by the use of a smart contract, which is a key element in the processing and presentation module, as it is responsible for demarcating and verifying users' access rights to information.

Considering these two flows, the smart contract becomes a crucial part of the whole system, as it is responsible for whole data access control. That's why the usage of a hybrid blockchain is necessary to provide clarity of algorithms used in that smart contract for every user of the proposed system. This will increase the trust in the network and opportunities for finding any vulnerabilities by users of the network.

Moreover, the usage of the right consensus mechanisms is crucial to the smooth working of such a system. Considering that the proposed system should process several transactions the consensus mechanism with high productivity should be chosen. In this case such mechanisms as proof of stake or delegated proof of stake.

Table 2 provides a comparison of the proposed architecture based on blockchain, smart contracts, and IPFS with traditional centralized systems, and public and private blockchains. Each row of the table discusses the advantages and disadvantages of each architecture in terms of data immutability, consent management, anonymization, decentralized storage, scalability, speed of access, legal aspects, data erasure, privacy and security, access control, transparency of data processing mechanisms, and auditing.

The proposed architecture, which combines blockchain, smart contracts, and IPFS, has significant advantages over traditional centralized systems, as well as public and private blockchains.

First, the high immutability and transparency of records guaranteed by using blockchain, makes this architecture more reliable compared to centralized systems where data can be changed or deleted. Consent management using smart contracts automates processes, reducing the risk of errors and delays that often occur in centralized systems.

Anonymization and pseudo-anonymization technologies allow for a high level of protection of personal information, which makes this architecture more reliable compared to centralized systems. Decentralized data storage with IPFS increases data availability and security, avoiding the single point of failure typical of centralized systems.

Access control implemented through smart contracts provides flexible and reliable protection of confidential information. Transparency of data processing mechanisms, achieved through smart contracts, increases user trust and simplifies auditing. Immutable records and transparent smart contracts also facilitate auditing and regulatory compliance, which is an important advantage in today's regulatory environment.

Table 2
Advantages and disadvantages of the proposed architecture for storing and processing personal data: comparison of blockchain and other systems

Criterion	The proposed architecture	Traditional centralized systems	Public blockchains	Private blockchains
Immutability and transparency of records	Advantages: Guarantee of immutability and transparency of records. Disadvantages: High complexity of implementation.	Advantages: Easy to change or delete records. Disadvantages: Low transparency and security.	Advantages: High transparency and immutability of records. Disadvantages: Less privacy.	Advantages: Controlled access to records. Disadvantages: Less transparency compared to public blockchains.
Consent management	Advantages: Automation of processes using smart contracts. Disadvantages: Difficulty setting up smart contracts.	Advantages: Ease of implementation. Disadvantages: High risk of errors and delays.	Advantages: Automated consent management. Disadvantages: High complexity of implementation.	Advantages: Automated consent management, easier to monitor. Disadvantages: Difficulty of implementation.
Anonymization and pseudo-anonymization	Advantages: Data privacy protection. Disadvantages: Requires additional resources for implementation.	Advantages: Ease of implementation. Disadvantages: High risk of data leakage.	Advantages: High level of anonymity. Disadvantages: Implementation difficulties for sensitive data.	Advantages: High confidentiality. Disadvantages: Less anonymous compared to public blockchains.
Decentralized data storage	Advantages: Increased availability and data security thanks to IPFS. Disadvantages: Complexity of implementation and management.	Advantages: Ease of management. Disadvantages: Single point of failure and high vulnerability to cyber-attacks.	Advantages: High reliability and availability. Disadvantages: High maintenance costs.	Advantages: High reliability and availability, easier to control. Disadvantages: High maintenance costs.
Scalability	Disadvantages: Scalability issues due to limits on the number of transactions.	Advantages: Easy scalability by adding servers. Disadvantage: Single point of failure risks.	Disadvantages: Scalability issues due to a large number of transactions.	Advantages: Better scalability compared to public blockchains. Disadvantages: There are scalability issues.
Data access speed	Advantages: Fast read speed Disadvantages: Slower write speed due to transaction confirmation and decentralized storage.	Advantages: High-speed data access.	Advantages: Fast read speed Disadvantages: Slower write speed due to transaction confirmation and decentralized storage.	Advantages: High-speed data access. Disadvantages: Possible delays due to internal checks.
Data deletion	Advantages: IPFS allows data deletion. Disadvantages: Data in the blockchain remains unchanged.	Advantages: Easy data deletion to meet GDPR requirements. Disadvantages: Low transparency and security.	Disadvantages: Data remains unchanged, making GDPR compliance difficult to achieve.	Disadvantages: Data remains unchanged, making GDPR compliance difficult to achieve.
Privacy and security	Advantages: High privacy and security thanks to hybrid blockchain, smart contracts, and IPFS. Disadvantage: High complexity of implementation.	Advantages: Ease of implementation. Disadvantage: Low privacy and security.	Advantages: High security. Disadvantage: Privacy may be compromised due to public access.	Advantages: High privacy and security. Disadvantage: Less transparency compared to public blockchains.
Access control	Advantages: Smart contracts provide flexible data access control. Disadvantage: Complexity in the smart contract development.	Advantages: Ease of access management. Disadvantage: High risk of unauthorized access.	Disadvantage: Open data access can be problematic for sensitive data.	Advantages: Flexible access control. Disadvantage: High setup and management costs.
Data processing mechanisms transparency	Advantages: Smart contracts ensure transparency of data processing. Disadvantages: High complexity of implementation.	Advantages: Ease of implementation. Disadvantages: Low transparency of data processing mechanisms.	Advantages: High transparency of data processing. Disadvantages: Privacy may be compromised.	Disadvantages: Less transparent compared to public blockchains.
Audit and compliance control	Advantages: Easily auditable due to immutable records and transparent smart contracts Disadvantages: High complexity of implementation.	Advantages: Ease of auditing. Disadvantages: Difficult to audit due to the possibility of changing or deleting records.	Advantages: Easy to audit due to immutable records. Disadvantages: Privacy may be compromised.	Advantages: Easy to audit. Disadvantages: Less transparent compared to public blockchains.

Therefore, the proposed architecture is more reliable, transparent, and secure compared to traditional centralized systems, and has advantages over public and private blockchains due to the possibility of flexible access management and data privacy protection. Although the implementation of such a system can be complex and costly. One of the methods that can help with the development of such a system can become the usage of dataflow design principles proposed in this study [20], as they will help to classify and represent information that is processed in the system which is an important part of access delimiting. Considering the requirements of the proposed system, the Solana network can be used as a basis for such a system, as it focuses on solving the scalability and speed issues of blockchain while providing an opportunity to create smart contracts and decentralized applications. The study [21] shows that the Solana network can be used in cases, where high speed of transactions and security are required.

Considering the advantages of the proposed system, they make it a promising solution for storing and processing personal data in today's world.

3. Conclusions

In this work, the concept of a service for collecting and storing personal data that meets the requirements of the General Data Protection Regulation (GDPR) was developed. The proposed service architecture uses innovative technologies, including blockchain, smart contracts, NFT, and the IPFS decentralized file system, to ensure a high level of privacy and data security.

The core components of the system include modules for registration and MFA, data collection and processing, as well as blockchain and IPFS data storage. The use of blockchain technology ensures the immutability and transparency of records of transactions and user consent, while IPFS guarantees the integrity and security of stored files.

In addition, the integration of smart contracts automates the management of consents and the implementation of privacy policies, and the mirror database allows you to work efficiently with anonymized data, keeping it safe and reducing the risk of disclosure of confidential information.

The integration of MFA increases the level of security at all stages of working with the system. MFA includes the use of passwords, SMS codes, authenticator applications, and biometric data, which provide an additional level of protection and prevent unauthorized access to personal data.

This approach allows to ensure compliance with GDPR requirements and increases user trust in the service. The use of multi-factor authentication ensures that only authorized users have access to the system, reducing the risk of unauthorized access and increasing the overall level of security.

In general, the proposed concept demonstrates the possibility of creating a reliable and effective system for collecting and storing personal data that meets modern requirements for protecting privacy and information security. The use of blockchain technologies, smart contracts, decentralized storage systems, and multi-factor authentication allows you to significantly improve data management and ensure their protection at a high level.

References

[1] F. Casino, T. Dasaklis, C. Patsakis, A Systematic Literature Review of Blockchain-Based Applications: Current Status, Classification and Open Issues, Telematics and Informatics 36 (2019) 55–81. doi: 10.1016/j.tele.2018.11.006.

[2] Y. Cheng, W. Jiang, Sharding and Layer 2 on Blockchain: A Comprehensive Survey, arXiv: preprint (2019).

[3] K. Christidis, M. Devetsikiotis, Blockchains and Smart Contracts for the Internet of Things, IEEE Access 4 (2019) 2292–2303. doi: 10.1109/ACCESS.2016. 2566339.

[4] M. Dong, et al., Blockchain for Secure and Trustworthy IoT: A Survey, IEEE Access 9 (2021) 4955–4971. doi: 10.1109/JIOT.2019.2920987.

[5] N. Kaaniche, M. Laurent, H. Ayed, A Blockchain-Based Data Usage Auditing Architecture with Enhanced Privacy Protection, IEEE Transactions on Dependable and Secure Computing (2020). doi: 10.1109/TDSC.2020.2965 094.

[6] L. Caldas, et al., Development of a Social Network for Research Support and Individual Well-Being Improvement, in: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2018) 383–386. doi: 10.1109/ASONAM.2018.8508365.

[7] X. Zhang, et al., Research on Blockchain Consensus Algorithm for Large-Scale High-Concurrency Power Transactions, 9th International Forum on Electrical Engineering and Automation (2022) 1221–1225. doi: 10.1109/IFEEA57288. 2022.10037907.

[8] B. Bèatrix-May, et al., Integrating Blockchain Technology in a Cyber Physical System to Secure Data, 24th International Conference on Control Systems and Computer Science (CSCS), (2023) 325–328. doi: 10.1109/CSCS592 11.2023.00058.

[9] C. Fan, R. Choubey, O. Rana, Blockchain-based Smart Contracts for IoT-based Autonomous Industrial Systems, in: IEEE Internet of Things Journal (2020). doi: 10.1109/JIOT.2020.2967715.

[10] S. Singh, N. Rajput, N. Kumar, A Decentralized Framework for Identity Management Using Blockchain, IEEE Access 8 (2020) 171653–171665. doi: 10.1109/ACCESS.2020.3024323.

[11] Y. Hu, J. Ni, A Blockchain-based Solution for Enhancing Security and Privacy in Online Social Networks, IEEE Transactions on Computational Social Systems (2021) doi: 10.1109/TCSS. 2021.3063921.

[12] M. Hasan, K. Salah, R. Jayaraman, Blockchain-Based Provenance for the Internet of Things: Opportunities, Challenges, and Directions, IEEE Internet of Things J. 7(5) (2020) 4039–4062. doi: 10.1109/JIOT.2019.2956522.

[13] M. Singh, P. Singh, T. Kim, Blockchain: A Game Changer for Securing IoT Data, Future Generation Comput. Syst. 139 (2023) 10–24. doi: 10.1016/j.future.202 3.01.005.

[14] S. Chen, X. Wang, J. Ren, Blockchain-Based Decentralized Identity Management for Secure Personal Data Sharing, Comput. Secur. 124 (2023). doi: 10.1016/j.cose.2023.103013.
[15] V. Poberezhnyk, V. Balatska, I. Opirskyy, Development of the Learning Management System Concept based on Blockchain Technology, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3550 (2023) 143–156.
[16] V. Balatska, et al., Blockchain Application Concept in SSO Technology Context, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654 (2024) 38–49.
[17] N. Al-Khater, M. Omar, R. Al-Debsi, Leveraging Blockchain for Privacy-preserving Data Aggregation in Smart Grids, Energy Reports 10 (2024) 202–214. doi: 10.1016/j.egyr.2023.12.004.
[18] S. Lee, D. Kim, Y. Park, Blockchain-based Secure Framework for IoT Device Authentication and Data Integrity, Internet of Things 22 (2023). doi: 10.1016/j.iot.2023.100526.
[19] T. Zhou, Y. Wang, H. Zhang, Blockchain-enabled Federated Learning for Secure and Privacy-preserving Data Sharing in Smart Cities, Future Generation Comput. Syst. 141 (2024) 98–109. doi: 10.1016/j.future.2023.02.012.
[20] O. Solomentsev, M. Zaliskyi, Method of sequential estimation of statistical distribution parameters in control systems design, IEEE 3rd International Conference on Methods and Systems of Navigation and Motion Control, MSNMC 2014 – Proceedings (2014) 135–138.
[21] Y. Averyanova, et al., UAS Cyber Security Hazards Analysis and Approach to Qualitative Assessment, Lecture Notes in Networks and Systems, 290 (2021) 258–265.
[22] S. Gnatyuk, Critical aviation information systems cybersecurity, Meeting Security Challenges Through Data Analytics and Decision Support (2016) 308–316.
[23] O. Deineka et al., Designing Data Classification and Secure Store Policy According to SOC 2 Type II, in: Cybersecurity Providing in Information and Telecommunication Systems, vol. 3654 (2024) 398–409.
[24] S. Vasylyshyn, I. Opirskyy, Combat Drone Swarm System (CDSS) Based on Solana Blockchain Technology, 7th International Workshop On Computer Modeling and Intelligent Systems, vol. 3702 (2024) 179–191