International Journal of Innovative Research in Engineering
Volume 5, Issue 3 (May-June 2024), PP: 33-37.
www.theijire.com
ISSN No: 2582-8746

Empower Data: A Privacy-Centric Consent-Based Sharing Solution

Koduru Anurup Reddy[1], Macha Shrenika[2], Kotha Tejasri[3], Lalu Banothu[4]
[1, 2,3] Department of CSE, Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.
[4]Associate Professor, Department of CSE, Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India.

How to cite this paper:
Koduru Anurup Reddy[1], Macha Shrenika[2], Kotha Tejasri[3], Lalu Banothu[4] "Empower Data: A Privacy-Centric Consent-Based Sharing Solution", IJIRE-V5I03-33-37.

Abstract: *Personal data is becoming increasingly valuable in business, as the insights that can be obtained from data processing continue to improve. However, it also can cause adverse effects on individuals. To improve data quality while satisfying privacy compliance, companies now have focused on collecting informed consent from individuals to directly handle personal data applying privacy-preserving techniques. Even though the companies obtain consent to use personal data, to improve transparency and accountability to ensure that companies deal with personal data according to consent, it is necessary for a system to comply with privacy requirements. Therefore, this paper proposes a new consent-based privacy-compliant personal data-sharing system that considers personal data-sharing flows and requirements obtained from enterprises and privacy frameworks, respectively. By analyzing a general process and the roles of actors for data sharing in enterprise environments according to standard privacy frameworks, this paper has proposed system requirements, system architecture, and detailed procedure for a consent-based privacy compliant processing method that considers compliance checking as well as consent checking. To show the feasibility of the proposed system, this paper demonstrates a prototype and the performance analysis in the lab and real-world environments.*

Keyword*: Data-sharing system, consent-based privacy, concept and framework.*

I.INTRODUCTION

Data is becoming increasingly valuable in business, as the insights that can be obtained from data processing continue to improve. Advancements in artificial intelligence and data processing technology [1] have enabled data-driven insights, uncovering business potentials and opportunities. Companies are now actively willing to utilize data to operate and expand their businesses. According to a report on big data analytics market size [2], the global market size is projected to grow from 307 Billion US dollars in 2023 to 745 Billion US dollars in 2030. Therefore, the importance of having more data has led companies to not only collect but also actively share and trade data (especially, personal data) among business stakeholders [3]. By combining and synthesizing large amounts of high-quality data, companies can gain deeper insights and improve their predictive capabilities. In this aspect, among all types of data, personal data plays a key role in maximizing the value of data in business by giving a basis for understanding and predicting customers' behavior as well as market trends. Thus, companies are making efforts to gather more personal data for their business through various channels. A company's proactive sharing and utilization of personal data can benefit its own business growth. However, this also can cause adverse effects on data providers (or data subjects) such as privacy infringement, unwanted marketing, and alleviated risk of data breaches. Since it is challenging to have data-driven innovation while protecting privacy [4], the necessity of guidance that can mitigate negative impacts and risks, safeguard data subjects' rights, and enable corporate data utilization has increased.

In response to these issues, many governments have implemented legal and regulatory frameworks, including the General Data Protection Regulation (GDPR) in the European Union [5], California Consumer Privacy Act (CCPA) in the United States [6], etc. These laws and regulations aim to protect individuals' personal data by setting rules and guidelines for companies and organizations to collect, process, store, and share personal data. Companies are now obliged to have systems and procedures in place for the legitimate use of personal data. Under the advent of new or more stringent regulatory frameworks, one approach to securely use and share (personal) data is applying privacy-preserving techniques. Applying privacy-preserving technologies (e.g., data anonymization, differential privacy, secure multi-

party computation, homomorphic encryption, etc.) [7], [8], [9] can make the sharing parties unable to recognize any personal data from the shared data so that the privacy regulations are no longer applied. By making personal data unidentifiable among the data processing parties, privacy-preserving technology ensures the protection of sensitive or personal data from unauthorized access, disclosure, or misuse. However, such technologies require extra data processing resources and can diminish the value of the data due to the loss of information in quantity and quality. Therefore, it is necessary to handle personal data without privacy- preserving techniques (i.e., handling personal data as it is for obtaining better quality information) in a privacy-compliant manner that supports and protects individuals' rights. To use and share personal data without compromising its quality and quantity while adhering to privacy regulations, it is necessary to consider various privacy-related compliance requirements for companies. Therefore, it is important to build a personal data-sharing system for supporting the data utilization stakeholders, which considers individuals' consent and other privacy compliance requirements. There are several ways to follow privacy compliance requirements (e.g., consent- based, privacy-preserving based, legal-based, etc.); particularly, considering consent-based personal data handling mechanisms draws attention from both academia and industry since new data-related regulations and governance has emphasized individuals' rights and consent management.

Personal data-sharing system needs to be implemented to cover the data utilization processes among the data subjects (or data providers), the data controller, the data processor, and the third party (or data requester) according to standards and regulations [5], [10]. Note that, in this paper, the terms ''data subject'' and ''data provider'' are used interchangeably, which means that an individual provides personal data to the data controller and the data processor. Figure 1 (inspired by [10], [11]) shows a general sequence of consent-based privacy-compliant personal data utilization among the actors. The company, which is responsible for providing a service to data subjects, acts as a data controller (usually acts as a data processor, too) and obtains personal data with necessary consent from the data subjects (or data providers). While the data controller manages the collected personal data and consent from data providers, it receives many data- sharing requests from data requesters who want to utilize personal data for their own purposes (e.g., internal departments, contracted third parties, or regulatory authorities, etc.). Upon receiving data-sharing requests from data requesters, the data controller assesses the requests.

If a data request is acceptable, the data controller instructs the data processor to process the requested dataset in accordance with the obtained consent, applicable requirements, and compliance. After the data processor reports the processing result, the data controller reviews and examines the processed dataset to make a decision. When the data controller decides to share, then the data requester receives a valid dataset.

II.RELATED WORKS

The objective of to use and share personal data without compromising its quality and quantity while adhering to privacy regulations, it is necessary to consider various privacy-related n other words, if companies can collect informed consent from individuals with proper purposes of personal data utilization and sharing, they are able to directly handle personal data without applying any privacy-preserving techniques.

This paper proposes a process for a privacy-compliant personal data-sharing system that should be considered for having transparency and accountability of personal data use of the companies according to the standard privacy, which consists of three actors for personal data utilization: the data requester, the data controller, and the data processor. Note that the informed consent and raw data are already collected and stored on the data controller side.

A privacy-preserving approach is a typical way to share data, which deletes/masks sensitive data using various techniques such as anonymization, pseudonymization, de-identification, etc. As a result, the shared data contains only non-sensitive data, and data providers are not deeply concerned about privacy infringement issues. There are several survey papers that introduced various privacy-preserving big data management and exchange models. One approach is applying the privacy-preserving technique from the data provider side. In other words, only the processed/filtered datasets can be collected from the data subjects (or data providers).

III.LITERATURE SURVEY

K. D. C. Adje, A. B. Letaifa, M. Haddad, and O. Habachi, open data are gold mines because they can be used to create services that develop a smart city while improving users' living conditions. Several research works go in this direction, presenting open data impact in the smart city for some, while others have focused on data processing methods. We have therefore deemed it necessary to make a state of the art on these different issues. The particularity of our study is that it shows the link between open data and smart city in all its aspects, describing what kind of open data is suitable for the smart city, how it is important for its development, and how these open data are processed to create services. Thus, in this article, we first present a review of existing surveys since 2015. Then, we present different smart city dimensions based on open data as well as some applications, and we detail how to process these data. We end with a list of open data sources as well as some challenges and solutions related to smart city services.

R. Podschwadt, D. Takabi, P. Hu, M. H. Rafiei, and Z. Cai, outsourced computation for neural networks allows users access to state-of-the-art models without investing in specialized hardware and know-how. The problem is that the users lose control over potentially privacy-sensitive data. With homomorphic encryption (HE), a third party can perform computation on encrypted data without revealing its content. In this paper, we reviewed scientific articles and publications in the particular area of Deep Learning Architectures for Privacy-Preserving Machine Learning (PPML) with Fully HE. We analyzed the changes to neural network models and architectures to make them compatible with HE and how these changes impact performance. Next, we find numerous challenges to HE-based privacy- preserving deep learning, such as computational overhead, usability, and limitations posed by the encryption schemes. Furthermore, we discuss potential solutions to the HE PPML challenges. Finally, we propose evaluation metrics that allow for a better and more meaningful comparison of PPML solutions.

N. Kim, H. Oh, and J. K. Choi, personal data have become the key to data-driven services and applications whereas privacy requirements are now strongly imposed by regulations. Meanwhile, people find it difficult to understand whether the services and applications handle personal data to comply with their agreements and regulations. Therefore, the need for privacy indicators, which summarize privacy contents as forms of privacy scoring, labels, etc., has increased to empower the users' rights by providing understandable information about privacy. For firm privacy indicators, proper criteria and methods for evaluating the level of privacy risks and compliance are required. Accordingly, this paper proposes a privacy scoring framework for services in the context of handling personal data, inspired by six standardized indicators. This paper introduces detailed information on standardized indicators and proposes privacy indicators to quantify privacy scores. Also, this paper proposes methods for evaluating privacy policy based on a set of machine learning-based hierarchical binary classifiers and processes for quantifying the level of privacy risks and compliance from privacy-related information. Through analyzing privacy policies and data access lists of more than 10,000 mobile applications on Google Play Store and investigating case studies on privacy scoring of some mobile applications, this paper shows the feasibility of the proposed framework.

Shojaee, Y. Zeng, M. Wahed, A. Seth, R. Jin, and I. Lourentzou, industrial Internet provides a collaborative computational platform for participating enterprises, allowing the collection of big data for machine learning tasks. Despite the promise of training and deployment acceleration, and the potential to optimize decision-making processes through data-sharing, the adoption of such technologies is impacted by the increasing concerns about information privacy. As enterprises prefer to keep data private, this limits interoperability. While prior work has largely explored privacy-preserving mechanisms, the proposed methods naively average or randomly sample data shared from all participants instead of selecting the most well-suited subsets for a particular downstream learning task. Motivated by the lack of effective data- sharing mechanisms for heterogeneous machine learning tasks in Industrial Internet, we propose PriED, a task-driven data-sharing framework that selectively fuses shared data and local data from participants to improve supervised learning performance. PriED utilizes privacy-preserving data distillation to facilitate data exchange, and dynamic data selection to optimize downstream machine learning tasks. We demonstrate performance improvements on a real semiconductor manufacturing case study.

G. Wu, S. Wang, Z. Ning, and B. Zhu, the digitization of Electronic Medical Record (EMR) provides potential access to a wealth of medical information, but also presents new challenges in privacy-preserved EMR exchanging and sharing. In this paper, we propose a blockchain-based smart healthcare system with fine- grained privacy protection for reliable data exchanging and sharing among different users. We design a blockchain-enabled dynamic access control framework combined with Local Differential Privacy (LDP) strategies to provide the attribute-based privacy protection in transaction workflow. We design four types of smart contracts in the framework to meet the requirements of anonymous transaction, dynamic access control, beneficial matching decision, and evaluation of published data in an open network. To satisfy fine-grained privacy protection, we classify sensitive attributes of EMRs into different levels and set differential privacy budgets to randomize attributes before data publishing. Also, we design data quality function to depict the disturbance incurred by LDP-based privacy preferences at the requester view, and present appropriate many-to-many matching decisions among participants for beneficial transactions. Finally, we develop a prototype system and test our approach using 200,000 real-world EMRs. Experimental results show that the proposed privacy-preserved scheme can make stable and reliable transactions between EMR publishers and requesters. The prototype system achieves individual-centric privacy configuration at the patient site, while providing error-guaranteed statistics at the requester site. Additionally, the access control policies, logs of anonymous transaction are kept in the block chain to provide system-level traceability.

Y. Tang, D. Gu, N. Ding, and H. Lu, anonymization technique has been extensively studied and widely applied for privacy- preserving data publishing. In most previous approaches, a microdata table consists of three categories of attributes, namely explicit-identifier, quasi-identifier (QI), and sensitive attribute. In general, individuals may have different views on the sensitivity of different attributes. Therefore, there is another type of attribute that contains both QI values and sensitive values, termed semi- sensitive attribute. In this paper, we propose a new anonymization technique, called Local Generalization and Bucketization, to prevent identity disclosure and protect the sensitive values on each semi-sensitive attribute and sensitive attribute. The rationale is to use local generalization and local bucketization to divide the tuples into local equivalence groups and partition the sensitive values into local buckets, respectively. The protections of local generalization and local bucketization are independent, so that they can be implemented by appropriate algorithms without weakening other protection. Besides, the protection of local bucketization for each semi-sensitive attribute and sensitive attribute is also independent. Consequently, local bucketization can comply with various principles in different attributes according to the actual requirements of anonymization. We conducted extensive experiments to illustrate the effectiveness of the proposed approach.

IV. PROPOSED WORK

Therefore, this paper proposes a new consent-based privacy-compliant personal data-sharing system that considers personal data-sharing flows and requirements obtained from enterprises and privacy, respectively. By analyzing a general process and the roles of actors for data sharing in enterprise environments according to standard privacy frameworks, this paper has proposed system requirements, system architecture, and detailed procedure for a consent-based privacy compliant processing method that considers compliance checking as well as consent checking. To show the feasibility of the proposed system, this paper demonstrates a prototype and the performance analysis in the real-world environments.

While the data controller manages the collected personal data and consent from data providers, it receives many data-sharing requests from data requesters who want to utilize personal data for their own purposes (e.g., internal departments, contracted third parties, or regulatory authorities, etc.). Upon receiving data-sharing requests from data requesters, the data controller assesses the requests. If a data request is acceptable, the data controller instructs the data processor to process the requested dataset in accordance with the obtained consent, applicable requirements, and compliance. After the data processor reports the processing result, the data controller reviews and examines the processed dataset to make a decision. When the data controller decides to share, then the data requester receives a valid database.

Consent Discovery and Matching: Data Storage Methods

This paper has mainly focused on text-based database stored in relational databases. Accordingly, the proposed consent based privacy-compliant processing method (Procedure 1) is for structured database store Moreover, consent discovery methods should be considered, which find relevant consent of data owners to match various raw data for processing personal data to generate shareable database for data requesters. Therefore, consent discovery and matching methods that can handle various different types of data storage methods in the futured in table- like databases. However, there are many different data storage methods.

Privacy-Preserving Techniques

Even though this paper proposes a consent-based personal data-sharing system that does not rely on privacy preserving methods, there are inevitable cases to apply privacy-preserving techniques to comply with certain privacy requirements regardless of the existence of the allowed consent from data providers.

Consent-Based Sharing in Block chain

As one approach, with the emergence of block chain technology, many studies have focused on utilizing the characteristics of the block chain (i.e., immutability, traceability, etc.) to check and enforce the consent of data providers for managing data access. Proposed a block chain-based consent management model for financial service platforms. The authors utilized a consortium block chain with a proof of authority mechanism to check both the informed consent of users and the certificates of regulators. Roman-Martinez et al. [18] suggested a service-oriented architecture for consent management, access control, and auditing of health data usage with a block chain. By utilizing two separate block chains for checking consent and auditing events, the authors developed a system with service-oriented architecture, which shares personal health data. However, block chain-based studies have performance issues (e.g., execution time, scalability, etc.) in applying real-world systems, particularly, for large-scale enterprise systems. Therefore, other approaches have considered mapping consent information into the datasets (or databases) within the existing systems. Particularly, some studies have focused on access management of datasets according to the consent information of data providers. In other words, data consumers are able to access only the consented dataset.

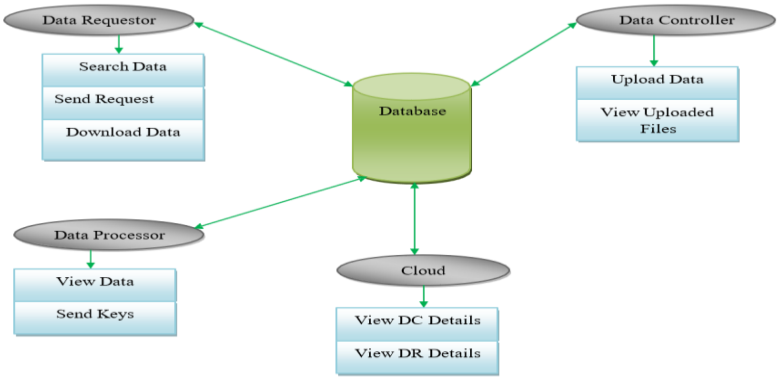

Figure 1. System architecture.

In this project data controller has a register with a user id and password. Data controller has a login with a user id and password. Data controller has a upload a data. Data requestor has a register with a user id and password. Data requestor has a search a data and send a request to the database. Data processor has a login with a user id and password. It was a view a data. and send a secret keys to copy to a key and send to a data requestor has got a keys then it was download a data. Cloud has a login with a user id and password. Cloud has a view data controller details and cloud has a data request details.

V.CONCLUSION

Since the issues of utilizing personal data while protecting privacy and data providers' right have focused, many companies now require tools for safely handling personal data. Especially, since identifying whether data is personal data or not becomes more difficult, a data provider's explicit consent on data utilization becomes more important to companies that want to utilize personal data. Therefore, this paper has proposed a consent-based privacy-compliant data sharing system. By analyzing a general process and the roles of actors for data-sharing in an enterprise environment, this paper has proposed system requirements that can support a consent-based privacy-compliant personal data-sharing system.

Future Enhancement

According to the identified requirements, this paper has proposed the system architecture and detailed procedure for a consent-based privacy-compliant processing method that considers

compliance checking as well as consent checking. For the demonstration, this paper also has presented a prototype implemented in a public cloud computing environment. Using the prototype, the performance analysis in the lab and real-world environments has shown that the proposed consent-based privacy-compliant personal data sharing system is feasible for real-world application.

References

1. *K. D. C. Adje, A. B. Letaifa, M. Haddad, and O. Habachi, ''Smart city based on open data: A survey,'' IEEE Access, vol. 11, pp. 56726–56748, 2023.*
2. *G. Malgieri and B. Custers, ''Pricing privacy—The right to know the value of your personal data,'' Comput. Law Secur. Rev., vol. 34, no. 2, pp. 289–303, Apr. 2018.*
3. *F. Schäfer, H. Gebauer, C. Gröger, O. Gassmann, and F. Wortmann, ''Datadriven business and data privacy: Challenges and measures for productbased companies,'' Bus. Horizons, vol. 66, no. 4, pp. 493–504, Jul. 2023.*

4. *General Data Protection Regulation (GDPR). Accessed: Jun. 13, 2023. [Online].*
5. *Ravindra Changala, "Evaluation and Analysis of Discovered Patterns Using Pattern Classification Methods in Text Mining" in ARPN Journal of Engineering and Applied Sciences, Volume 13, Issue 11, Pages 3706-3717 with ISSN:1819-6608 in June 2018.*
6. *Ravindra Changala "A Survey on Development of Pattern Evolving Model for Discovery of Patterns in Text Mining Using Data Mining Techniques" in Journal of Theoretical and Applied Information Technology, August 2017. Vol.95. No.16, ISSN: 1817-3195, pp.3974-3987.*
7. *California Consumer Privacy Act (CCPA). Accessed: Jun. 13, 2023. [Online]. Available:*
8. *J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, ''Ensuring security and privacy preservation for cloud data services,'' ACM Comput. Surv., vol. 49, no. 1, pp. 1–39, Mar. 2017.*
9. *F. N. Wirth, T. Meurers, M. Johns, and F. Prasser, ''Privacy-preserving data sharing infrastructures for medical research: Systematization and comparison,'' BMC Med. Informat. Decis. Making, vol. 21, no. 1, p. 242, Aug. 2021.*
10. *Ravindra Changala, Framework for Virtualized Network Functions (VNFs) in Cloud of Things Based on Network Traffic Services, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 Volume 11, Issue 11s, August 2023.*
11. *Ravindra Changala, Block Chain and Machine Learning Models to Evaluate Faults in the Smart Manufacturing System, International Journal of Scientific Research in Science and Technology, Volume 10, Issue 5, ISSN: 2395-6011, Page Number 247-255, September-October-2023.*
12. *R. Podschwadt, D. Takabi, P. Hu, M. H. Rafiei, and Z. Cai, ''A survey of deep learning architectures for privacy-preserving machine learning with fully homomorphic encryption,'' IEEE Access, vol. 10, pp. 117477–117500, 2022.*
13. *Information Technology—Security techniques—Privacy Framework, International Organization for Standardization, ISO Central Secretary, Geneva, Switzerland, Standard ISO/IEC 29100:2011, 2017. [Online]. Available:*
14. *N. Kim, H. Oh, and J. K. Choi, ''A privacy scoring framework: Automation of privacy compliance and risk evaluation with standard indicators,'' J. King Saud Univ., Comput. Inf. Sci., vol. 35, no. 1, pp. 514–525, Jan. 2023.*
15. *Ravindra Changala, "Optimization of Irrigation and Herbicides Using Artificial Intelligence in Agriculture", International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(3), pp. 503–518.*
16. *Ravindra Changala, AIML and Remote Sensing System Developing the Marketing Strategy of Organic Food by Choosing Healthy Food, International Journal of Scientific Research in Engineering and Management (IJSREM), Volume 07 Issue 09, ISSN: 2582-3930, September 2023.*
17. *J. Zhang and C. Dong, ''Privacy-preserving data aggregation scheme against deletion and tampering attacks from aggregators,'' J. King Saud Univ., Comput. Inf. Sci., vol. 35, no. 4, pp. 100–111, Apr. 2023.*
18. *P. Shojaee, Y. Zeng, M. Wahed, A. Seth, R. Jin, and I. Lourentzou, ''Task-driven privacy- preserving data-sharing framework for the industrial internet,'' in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2022, pp. 1505–1514.*
19. *Ravindra Changala, "Sentiment Analysis in Social Media Using Deep Learning Techniques", International Journal of Intelligent Systems and Applications in Engineering, 2024, 12(3), 1588–1597.*
20. *Ravindra Changala, "Integration of IoT and DNN Model to Support the Precision Crop", International Journal of Intelligent Systems and Applications in Engineering, Vol.12 No.16S (2024).*
21. *Ravindra Changala, "UI/UX Design for Online Learning Approach by Predictive Student Experience", 7th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2023 - Proceedings, 2023, pp. 794–799, IEEE Xplore.*
22. *Z. Xiao, X. Fu, and R. S. M. Goh, ''Data privacy-preserving automation architecture for industrial data exchange in smart cities,'' IEEE Trans. Ind. Informat., vol. 14, no. 6, pp. 2780– 2791, Jun. 2018.*
23. *G. Wu, S. Wang, Z. Ning, and B. Zhu, ''Privacy-preserved electronic medical record exchanging and sharing: A blockchain-based smart healthcare system,'' IEEE J. Biomed. Health Informat., vol. 26, no. 5, pp. 1917–1927, May 2022.*
24. *Ravindra Changala, Development of CNN Model to Avoid Food Spoiling Level, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, ISSN: 2456-3307, Volume 9, Issue 5, September-October-2023, Page Number 261-268.*
25. *B. Li and K. He, ''Local generalization and bucketization technique for personalized privacy preservation,'' J. King Saud Univ., Comput. Inf. Sci., vol. 35, no. 1, pp. 393–404, Jan. 2023.*