International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 10, Issue 1, pp 66-73, January-2026
https://doi.org/10.47001/IRJIET/2026.101008

Privacy-Preserving Record Linkage: A Survey of Key Concepts and Approaches

[1]Patel Krupali, [2]Dr. Prashant Pittalia

[1]Prof. V. B. Shah Institute of Management, R.V. Patel College of Commerce (English Medium), V. L. Shah College of Commerce (Gujarati Medium) and Sutex Bank College of Computer Applications & Science, Surat, Gujarat, India
[2]Department of Computer Science, Sardar Patel University, Vallabh Vidyanagar, Gujarat, India

***Abstract* - In today's era massive data sets having large and complex structure with the difficulties of storing, analysing and visualizing for further processes or results. The voluminous data, especially personal data in multiple sources, present large opportunities and insight for businesses for analysis and investigating the value of linked and integrated data. Privacy is a major concern while we share or link data through networks of different organizations. Privacy Preserving Record Linkage (PPRL) aims to address this problem by identifying and linking records that correspond to the same real world entity across several data sources held by different parties without revealing any sensitive information about these entities. Data deduplication is intelligent comparison or single instance storage. It is a process that eliminates redundant copies of data and reduces storage overhead. In this article, we provide an overview of the research literature in privacy-preserving record linkage, discuss the different types of techniques that have been proposed. We conclude this work with an overview of PPRL techniques.**

Keywords: Privacy, Data-Linkage, Record Linkage, Data-analysis, Data comparison.

I. Introduction

Privacy-preserving record linkage (PPRL) is a methodology that allows for linking person-level data from disparate sources while mitigating privacy concerns. [1] Big data is term for massive data sets having large and complex structure with the difficulties of storing, analysing and visualizing for further processes or results. The voluminous Big data especially personal data in multiple source, present large opportunities and insight for businesses for analysis and investigate the value of linked and integrated data. Privacy is major concern while we share or link data through network of different organization.

Privacy Preserving Record Linkage (PPRL) aims to address this problem by identifying and linking records that correspond to the same real world entity across several data source held by different parties without revealing any sensitive information about these entities. PPRL for Big Data poses several challenges are such as Scalability to multiple large databases, achieving high quality results of the linkage in presence of variety and velocity of data and preserving privacy and confidentiality of the entities represented in Big data collection.[30] Data deduplication is intelligent comparison or single instant storage. It is process that eliminates redundant copies of data and reduces storage overhead [2]

II. Motivation and Scope

A wide range of applications such as government services, healthcare, crime and fraud detection, national security, require personal identifiable data from multiple sources held by different organizations to be integrated or linked. Integrated data can then be used for data mining and analytics to empower efficient and quality decision making with rich data.[3] [4][5]

Public health research, Business collaborations, National security, Geocode matching to match the addresses in different database to geographic locations which allow the spatial analysis [31]. As these scenarios highlight, there is a need for techniques that allow data matching in such ways that the database owners do not have to reveal any of their sensitive private or confidential data to any other organization involved in a matching exercise, and only the matched records are being disclosed to the organization(s) or individual(s) that require them.[6]

Organizations keep their data tightly isolated from other systems, are major barrier to the effective use of data analytics in many fields. Unfortunately, when the data in question involves information about people, and their personal data such as name, surname, address it is necessity querying or joining based on personally identifiable data that could be used to explicitly identify an individual. It is important to protect personal data, both while it is at rest on system and shared among networks. These studies highlight the diversity of approaches in the field, each addressing specific concerns like security, scalability, real-time applicability, and comprehensive evaluation. Future research could bridge the gaps between these areas, optimizing for both security and efficiency, especially in large-scale or real-time environments.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 1, pp 66-73, January-2026

https://doi.org/10.47001/IRJIET/2026.101008

The dimensions used to characterize PPRL techniques. reviewed the literature related to string matching and secure string matching in the context of privacy-preserving record linkage (PPRL), the string matching is increasingly important in many domains.

III. Privacy Preserving Record Linkage Generic Framework

Privacy-preserving record linkage (PPRL) is a methodology that allows for linking person-level data from disparate sources while mitigating privacy concerns [7]. It ensures privacy by encrypting or masking PII used for matching [10].

PPRL Process

1. *Data Pre-processing:* This involves cleaning and standardizing data to ensure consistency and accuracy before the linkage process. Steps may include handling missing values, correcting errors, and formatting data uniformly.

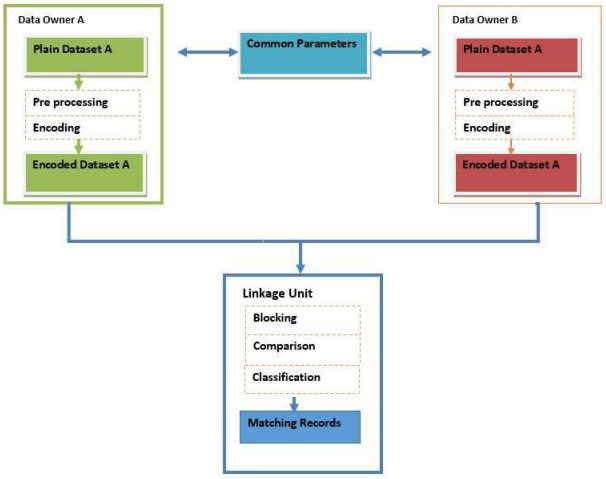

Figure 1: Generic Framework of PPRL

2. *Encoding:* Refers to the process of transforming sensitive data elements into a protected format, often by replacing original values with pseudo-values or encrypted versions to prevent unauthorized access.
3. *Blocking/Filtering:* Blocking: Blocking reduces the number of record comparisons by dividing data into smaller, more manageable blocks based on a common attribute or key. Only records within the same block are compared, which significantly improves efficiency.
4. *Filtering:* Filtering involves applying criteria to exclude records that are unlikely to match, further reducing the number of comparisons needed. This can involve pre-screening data to remove irrelevant records.
5. *Comparison:* This step involves comparing the attributes of records within the same block using various similarity measures (e.g., Jaro-Winkler, Levenshtein distance for strings) to determine if they are likely matches.
6. *Classification:* After comparing records, the results are classified into matches, non-matches, or potential matches. This can be done using deterministic rules or probabilistic models.

Applications of PPRL

Data Integration: Organizations often collect data from multiple sources or datasets for various purposes such as research, analysis, or service provision. PPRL enables these organizations to integrate datasets without compromising the privacy of individuals whose data is being merged.

Healthcare: In healthcare, there may be a need to link patient records across different hospitals or healthcare providers to get a comprehensive view of a patient's medical history. PPRL techniques ensure that this linkage can be performed securely without revealing sensitive patient information. PPRL enables researchers to link patient data from disparate sources (e.g., hospitals, registries) without compromising patient privacy. This facilitates large-scale studies on disease prevalence, treatment outcomes, and healthcare utilization. [8] Explores the challenges and opportunities in Australian health data linkage research.

Government Services: Government agencies often maintain separate databases containing information about citizens, such as tax records, social services records, or census data. PPRL allows these agencies to link records across datasets to improve service delivery or policy-making while protecting citizen privacy.

Research: Researchers may need to link datasets from different studies or sources to conduct comprehensive analyses. PPRL enables this linkage while safeguarding the privacy of individuals whose data is being used for research purposes.

Marketing and Business Intelligence: Companies may want to merge customer data from various sources to gain insights into consumer behavior or to provide better-targeted marketing campaigns. PPRL ensures that this data linkage can be performed without compromising customer privacy.

Law Enforcement and Criminal Justice: Law enforcement agencies may need to link records across various databases to investigate crimes or track individuals. PPRL techniques help ensure that this can be done securely without violating individuals' privacy rights.

Research Publication References: By employing PPRL techniques in systematic review searches, researchers can

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 1, pp 66-73, January-2026

https://doi.org/10.47001/IRJIET/2026.101008

effectively aggregate and analyses data from multiple sources while safeguarding the privacy of the individuals or organizations mentioned in those sources. This is crucial for maintaining ethical standards and ensuring compliance with data protection regulations.

Income Tax Department Information: Aims to identify and link records that correspond to the same entity or individual across different databases based on the matching of personal identifying attributes, such as name and address, without revealing the actual values in these attributes due to privacy concerns.

IV. Categorization of Data Matching Approaches

1. Deterministic Matching approaches

Deterministic privacy-preserving record linkage is an approach used to connect records from different databases while ensuring that sensitive personal information is protected from unauthorized access. This method utilizes specific algorithms that allow for the matching of records based on unique identifiers or features that remain consistent across databases, without disclosing identifiable information by[9] [8]. Each deterministic encryption scheme has its own set of considerations, and selecting the appropriate method depends on the specific privacy, efficiency, and data characteristics required for the record linkage process. While cryptographic hashing is the most widely used approach due to its simplicity and robust privacy protection, alternative techniques may be explored based on the complexity and demands of the system.

2. Probabilistic Matching approaches

Probabilistic privacy preserving record linkage is a method in data management that allows for the secure and anonymous merging of datasets while maintaining individual privacy. This approach typically involves probabilistic models to assess the likelihood that records from different sources pertain to the same individual, allowing for more accurate matching of records compared to deterministic methods [10]. Probabilistic encryption techniques such as Homomorphic Encryption (Partially Homomorphic Encryption), Order-Preserving Encryption (OPE), Fully Homomorphic Encryption (FHE), Randomized Hashing (with Salt), Searchable Encryption (with Probabilistic Methods), Public Key Encryption (with Randomization), Blinding Techniques play a crucial role in privacy-preserving record linkage by adding randomness, which prevents attackers from inferring patterns in encrypted data. These methods strike a balance between security and functionality, enabling the matching and linking of records across datasets while safeguarding sensitive information. The selection of a particular probabilistic encryption method depends on the specific needs of the PPRL application, such as computational efficiency, the nature of the data being matched, and the level of privacy protection required.

3. Combined Matching approaches

It is a hybrid approach that utilizes probabilistic techniques to gauge the likelihood of matches while employing deterministic methods to ensure a high degree of precision and recall. The exponential growth of patient data has created opportunities for healthcare research but also significant challenges in securely linking data across multiple sources. The study demonstrates that no single PPRL tool currently meets all practical needs. A hybrid approach leveraging the strengths of E-PIX (Encrypted Privacy Information Exchange). It is a tool designed for privacy preserving record linkage that uses advanced encryption and hardening techniques to secure data during the linkage process. And MainSEL (Mainzelliste SecureEpiLinker) It is an extension of the Mainzelliste pseudonymization service and employs secure multi-party computation (SMPC) for privacy-preserving record linkage, ensuring that sensitive data is not shared directly between parties could provide a robust and adaptable solution. The authors call for more research on interoperability and enhanced usability to optimize PPRL methodologies for diverse healthcare scenarios.[34] Combine traditional models like Fellegi-Sunter with modern methods like differential privacy or federated learning as hybrid approach cab be balance trade-offs between accuracy, scalability, and privacy.

Structured survey analysis table for the comparative study of literature reviews:

No	Publication Year	Reference	Tools & Techniques	Datasets	Algorithms	Limitations	Future Scope
1	2024	[11]	Python (3.7), Intel i7-8565U, Windows	Real and synthetic: NCVR datasets	Data Preparation and Generation (BF), Approximate Record Linkage, BF Dice Coefficient Similarity	Lack of optimization techniques to reduce time cost	Improve efficiency, linkage quality, and privacy for big data
2	2024	[12]	Python, Intel i7-10750H, Ubuntu 20.04	NCVR (100K–1M), EURO (25,343)	Block generation (DO), One-to-One and Many-to-One	One-to-one approach is slower but	Improve Bloom filter to reduce errors and

ISSN (online): 2581-3048
Volume 10, Issue 1, pp 66-73, January-2026
https://doi.org/10.47001/IRJIET/2026.101008

					encryption (DO, LU)	accurate; many-to-one is faster but less accurate	enhance time complexity
3	2024	[13]	PyCharm (2023.3)	NCVR dataset	HE-PPRL, RBF-PPRL, F-PPRL, MD-PPRL	Evaluations focus only on single indicators linkage, security, or efficiency)	Develop comprehensive frameworks balancing multiple metrics
4	2024	[14]	Python	Simulated and real-world datasets	Fellegi-Sunter model, DP Regression (Noisy Gradient Descent, Sufficient Statistics Perturbation)	Increased complexity and noise in DP algorithms	Extend DP methods to general supervised learning and improve privacy in federated learning
5	2024	[15]	Python	Clinical, vaccination records, health data	Noisy Gradient Descent, Sufficient Statistics Perturbation, Fellegi-Sunter model	Scalability, uncertainty in linkage results	Enhance scalability and efficiency; explore ethical considerations
6	2023	[16]	Python 2.7, Xeon 2.1 GHz, Ubuntu 18.04	NCVR, EURO datasets	BF Encoding, TMH Encoding, MMK Encoding, 2SH Encoding, SLK Encoding	Limited vulnerability considerations	Analyze the impact of database size on vulnerabilities
7	2023	[17]	Python 3.5.2	NCVR datasets	Dice coefficient, Privacy-preserving BF encoding with Local Differential Privacy	Cardinality estimation not robust to data errors	Develop typo and incomplete data models using deep learning for PPRL
8	2023	[18]	Python	NCVR, Ohio voter registration datasets	Cosine similarity (autoencoder), Bloom Filter, Autoencoder	Scalability	Explore different autoencoder configurations for improved scalability
9	2023	[19]	-	Real-world datasets	Linkage with attribute-level similarities, CLK-RBF encoding	Integration of Bloom filter hardening	Apply value-specific weight for improved encoding
10	2023	[NO_PRINTED_FORM]	Blockchain	Brazilian politician dataset	Splitting Bloom Filter, Blockchain-Based PPRL, Deep Learning classifiers	Slow performance; SBF error	Investigate SBF errors, integrate differential privacy
11	2023	-	Java 11+, Maven	Voter, music, consumer product datasets	Apriori-like approach, clusteringbased linkage	Lack of secured linkage methods	Enhance security and scalability through cluster-based approaches
12	2023	Han [NO_PRINTED_FORM] et al.	Python 3.6.5	NCVR dataset	BF-SNN Blocking Method (shortest nearest neighbourhood)	Efficiency improvements needed for big data	Improve efficiency for large-scale linkage
13	2023	[20]	Python 3.8.6, Intel i7-10750H, Ubuntu 20.04	NCVR dataset	Diffusion layer, Bloom Filter Encoding	Vulnerable to graph matching attacks and correlations	Explore non-linear diffusion, dynamic diffusion layers
14	2022	[21]	LinXmart	WA Hospital Morbidity, WA Death Registrations	Bloom Filter Encoding	Enhanced pre-processing techniques required	Standardize data formats, reduce data entry errors
15	2022	[22]	Python 2.7	Synthetic data (credit card, barcode, IBAN)	Encoded q-grams comparison, Bit array generation	Algorithm complexity and execution time	Explore scalability for real-world data

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 1, pp 66-73, January-2026

https://doi.org/10.47001/IRJIET/2026.101008

16	2022	[23]	Java 64-bit, Windows 10	Real and synthetic datasets	NameGist, K-Anonymization, Naive incremental clustering	Low linkage accuracy	Improve linkage quality with better algorithms
17	2022	[24]	Python 3.7	Training datasets (various domains)	Local model generation, Classification by LU	Scalability to large databases	Enable real-time counting and dynamic updates without reclustering
18	2022	[8]	-	Healthcare dataset	Deterministic Approach by Bloom Filter. Dice coefficients, and Yao's garbled circuits, blocking and partitioned Bloom filters.	Handling data error	Improving Error Tolerance, scalability
19	2021	[25]	Blockchain	Distinct datasets	3PAC	Slow performance; SBF error	Investigate SBF error, integrate differential privacy
20	2021	[26]	Python 2.7	NCVR (varied string types)	Suffix Tree Encoding, Secure First Character Encoding	-	-
21	2020	[27]	C#, Windows 7	NCVR	Soundex Encoding, Bloom Filter Encoding	Scalability and security	Enhance Bloom filter for security
22	2018	[29]	Java	Synthetic dataset, (http://www.mockaroo.com)	Algorithm 1 Computing/ Comparing Dice Coefficient of Bloom Filters	High computational cost of using, not easily scalable for Very large datasets.	Efficient by parallelizing the process, test the method on real-world datasets
23	2017	[28]	Java	Synthetic dataset (Car Insurance)	Phonetic algorithm, Blocking Algorithms	Need for faster and secure techniques	Develop scalable PPRL for numeric values

Year-wise summary of the most frequently used databases, along with their usage counts:

Year	Most Used Database	Usage Count
2024	NCVR datasets	3
2023	NCVR datasets	4
2022	Synthetic data (credit card, barcode, IBAN)	2
2021	NCVR (varied string types)	1
2020	NCVR	1
2018	Synthetic dataset (http://www.mockaroo.com)	1
2017	Synthetic dataset (Car Insurance)	1

Comparison Summary Table:

Aspect	Early Approaches	Recent Advancements
Tools	C#, Java	Python, Blockchain Technology
Datasets	Single real-world datasets (e.g., NCVR)	Real and synthetic datasets from multiple domains
Algorithms	Basic Bloom Filter encoding	Advanced privacy mechanisms (Differential Privacy, Hardened BF, Autoencoders)
Limitations	Scalability, accuracy	Scalability, typo handling, security vulnerabilities
Future Scope	Enhance efficiency and security	Big data readiness, deep learning applications, cross-domain implementation

ISSN (online): 2581-3048
Volume 10, Issue 1, pp 66-73, January-2026
https://doi.org/10.47001/IRJIET/2026.101008

V. Key Challenges in Privacy-Preserving Record Linkage (PPRL)

Privacy-preserving record linkage (PPRL) plays a crucial role in ensuring data privacy while linking records across different datasets. Despite its promising capabilities, several key challenges persist in the field:

1. Scalability and Efficiency

Handling large-scale datasets while maintaining privacy and efficiency remains a significant challenge. As datasets grow in size and complexity, existing algorithms can struggle with processing time and memory usage. As an impact the volume of data increases, methods need to scale without compromising on the quality of linkage or the level of privacy preservation.[15]

3. Security Vulnerabilities

Ensuring data privacy and security while performing record linkage can be difficult. Common attacks such as graph matching attacks or correlation attacks can undermine the security of PPRL systems. If the privacy of sensitive data is compromised, it could result in privacy breaches or unauthorized access to personal information.[20] [19]

4. Algorithm Complexity and Runtime

Many advanced PPRL algorithms, especially those designed to maintain high security and privacy, are computationally expensive. These can result in slow performance, especially when handling large datasets or when real-time linkage is required. [29]

5. Data Privacy and Differential Privacy

Achieving a balance between privacy and utility is a significant issue. [30] [12], [31]

6. Heterogeneous Data

As a challenge Different datasets often contain heterogeneous data formats, such as structured, semi-structured, and unstructured data, which complicate the linkage process. [19] Often face challenges with data pre-processing when linking heterogeneous datasets, such as combining healthcare data with voter registration records.

7. Integration with Federated Learning

Federated learning, where data remains decentralized, is increasingly being explored for privacy preserving record linkage is challenging. However, combining federated learning with PPRL introduces new challenges related to data synchronization, communication overhead, and maintaining privacy across distributed nodes.[30] discussed the integration of Differential Privacy (DP), but issues with scalability and data synchronization in a federated context remain largely unsolved.

8. Ethical and Legal Considerations

The ethical and legal implications of PPRL, especially when handling sensitive data such as medical records, voting data, or financial information, are a major concern. [15] Called for exploring ethical considerations in public health data, particularly when dealing with personal health information.

VI. Future Directions for Privacy-Preserving Record Linkage (PPRL)

Current research in privacy-preserving record linkage (PPRL) reveals gaps and opportunities for future exploration, such as improving scalability, developing hybrid privacy models, and enhancing automation to address growing data volumes and complexity. [32]. [33] Emerging technologies like AI, machine learning, and block-chain introduce both opportunities and new privacy threats, requiring strategies to address evolving adversarial attacks. Advances in cryptography, including quantum cryptography, and innovative privacy-preserving models offer promising avenues to strengthen security and efficiency. Additionally, there is a pressing need for interoperability and standardized PPRL protocols across industries to ensure consistency and broader applicability. Ethical and regulatory challenges remain central to PPRL, as balancing individual privacy with the social and economic benefits of data linkage continues to be a critical concern, demanding frameworks that prioritize both fairness and compliance.

VII. Conclusion

The key challenges in PPRL revolve around balancing privacy, security, scalability, and accuracy while addressing complex data types and legal constraints. Future research should focus on developing efficient algorithms, improving security against advanced threats, enhancing scalability for real-time applications, and creating frameworks that can handle heterogeneous data efficiently. At the same time, ethical and legal considerations must be integrated into the development of these privacy-preserving methods.

REFERENCES

[1] S. J. Petersen, R. D. Lieberthal, K. J. Miller, and N. H. Vakil, “Privacy Preserving Record Linkage (PPRL) Strategy and Recommendations Sponsor: National Institute on Aging PPRL Linkage Strategies Report

ISSN (online): 2581-3048
Volume 10, Issue 1, pp 66-73, January-2026
https://doi.org/10.47001/IRJIET/2026.101008

McLean, *VA*," 2023. [Online]. Available: https://www.alz.org/alzheimers-dementia/facts-figures.

[2] M. Ostermann, I. Nesterow, and M. Wolfien, "A Hybrid-Approach for Privacy Preserving Record Linkage - A Case Study from Germany," *Stud Health Technol Inform*, vol. 316, pp. 43–47, Aug. 2024, doi: 10.3233/SHTI240340.

[3] D. Vatsalan, P. Christen, C. O'keefe, and V. S. Verykios, "An Evaluation Framework for Privacy-Preserving Record Linkage," 2014. [Online]. Available: http://repository.cmu.edu/jpc

[4] P. Christen, "Privacy-Preserving Data Linkage and Geocoding: Current Approaches and Research Directions."

[5] T. Churches and P. Christen, "Some methods for blindfolded record linkage," *BMC Med Inform DecisMak*, vol. 4, Jun. 2004, doi: 10.1186/1472-6947-4-9.

[6] M. Franke, V. Christen, P. Christen, F. Rohde, and E. Rahm, "(Privately) Estimating Linkage Quality for Record Linkage," *in Advances in Database Technology - EDBT, OpenProceedings.org*, Nov. 2023, pp. 294–306. doi: 10.48786/edbt.2024.26.

[7] A.P. Brown, A. M. Ferrante, S. M. Randall, J. H. Boyd, and J. B. Semmens, "Ensuring privacy when integrating patient-based datasets: New methods and developments in record linkage," *Front Public Health*, vol. 5, no. MAR, Mar. 2017, doi: 10.3389/FPUBH.2017.00034.

[8] S. B. T. G. , B. B. , S. G. E. , T. M. , E. T. , K. N. S. L. , N. C. , M. A. N. T. , D. L. and R. H. S. Alisia Southwell1, "2022-Validating a novel deterministic privacy-preserving record linkage between," *Int J Popul Data Sci*, Nov. 2022.

[9] T. Ranbaduge, D. Vatsalan, and M. Ding, "Privacy-preserving Deep Learning based Record Linkage," *IEEE Trans Knowl Data Eng*, 2023, doi: 10.1109/TKDE.2023.3342757.

[10] K. Schmidlin, K. M. Clough-Gorr, and A. Spoerri, "Privacy Preserving Probabilistic Record Linkage (P3RL): A novel method for linking existing health-related data and maintaining participant confidentiality," *BMC Med Res Methodol*, vol. 15, no. 1, May 2015, doi: 10.1186/s12874-015-0038-6.

[11] S. Han, Z. Wang, D. Shen, and C. Wang, "A Parallel Multi-Party Privacy-Preserving Record Linkage Method Based on a Consortium Blockchain," *Mathematics*, vol. 12, no. 12, p. 1854, Jun. 2024, doi: 10.3390/math12121854.

[12] S. Vaiwsri, T. Ranbaduge, and P. Christen, "Encryption-based sub-string matching for privacy preserving record linkage," 2024.

[13] S. Han, Y. Wang, D. Shen, and C. Wang, "A Multi-Party Privacy-Preserving Record Linkage Method Based on Secondary Encoding," *Mathematics*, vol. 12, no. 12, p. 1800, Jun. 2024, doi: 10.3390/math12121800.

[14] S. Lin, E. Kolaczyk, and E. D. Kolaczyk, "NBER WORKING PAPER SERIES DATA PRIVACY FOR RECORD LINKAGE AND BEYOND Data Privacy for Record Linkage and Beyond," 2024. [Online]. Available: http://www.nber.org/papers/w32940.

[15] A.Pathak et al., "Privacy preserving record linkage for public health action: opportunities and challenges," *Journal of the American Medical Informatics Association*, Nov. 2024, doi: 10.1093/jamia/ocae196.

[16] A.Vidanage, P. Christen, T. Ranbaduge, and R. Schnell, "A Vulnerability Assessment Framework for Privacy-preserving Record Linkage," *ACM Transactions on Privacy and Security*, vol. 26, no. 3, Jun. 2023, doi: 10.1145/3589641.

[17] N. Wu, D. Vatsalan, M. A. Kaafar, and S. K. Ramesh, "Privacy-Preserving Record Linkage for Cardinality Counting," Jan. 2023, [Online]. Available: http://arxiv.org/abs/2301.04000.

[18] V. Christen, T. Häntschel, P. Christen, and E. Rahm, "Privacy-preserving record linkage using autoencoders," *Int J Data Sci Anal*, vol. 15, no. 4, pp. 347–357, May 2023, doi: 10.1007/s41060-022-00377-2.

[19] T. Nóbrega, C. Eduardo S. Pires, and D. Cassimiro Nascimento, "Towards Auditable and Intelligent Privacy-Preserving Record Linkage," *Sociedade Brasileira de Computacao - SB*, Oct. 2023, pp. 270–284. doi: 10.5753/sbbd_estendido.2023.232442.

[20] F. Armknecht, Y. Heng, and R. Schnell, "Strengthening Privacy-Preserving Record Linkage using Diffusion," *Proceedings on Privacy Enhancing Technologies*, vol. 2023, no. 2, pp. 298–311, Apr. 2023, doi: 10.56553/popets-2023-0054.

[21] S. Randall et al., "A blinded evaluation of privacy preserving record linkage with Bloom filters," *BMC Med Res Methodol*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12874-022-01510-2.

[22] S. Vaiwsri, T. Ranbaduge, and P. Christen, "Accurate and efficient privacy-preserving string matching," *Int J Data Sci Anal*, vol. 14, no. 2, pp. 191–215, Aug. 2022, doi: 10.1007/s41060-022-00320-5.

[23] S. I. Khan, A. B. A. Khan, and A. S. M. L. Hoque, "Privacy preserved incremental record linkage," *J Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00655-7.

[24] S. Vaiwsri, T. Ranbaduge, and P. Christen, "Accurate and efficient privacy-preserving string matching," *Int J*

ISSN (online): 2581-3048
Volume 10, Issue 1, pp 66-73, January-2026
https://doi.org/10.47001/IRJIET/2026.101008

Data Sci Anal, vol. 14, no. 2, pp. 191–215, Aug. 2022, doi: 10.1007/s41060-022-00320-5.

[25] T. Nóbrega, C. E. S. Pires, and D. C. Nascimento, "Blockchain-based Privacy-Preserving Record Linkage: enhancing data privacy in an untrusted environment," *Inf Syst*, vol. 102, Dec. 2021, doi: 10.1016/j.is.2021.101826.

[26] S. Vaiwsri, T. Ranbaduge, P. Christen, and K. S. Ng, "Accurate and Efficient Suffix Tree Based Privacy-Preserving String Matching," Apr. 2021, [Online]. Available: http://arxiv.org/abs/2104.03018.

[27] N. Shekokar and V. M. Shelake, "An Enhanced Approach for Privacy Preserving Record Linkage during Data Integration," *in 2020 6th IEEE International Conference on Information Management, ICIM 2020, Institute of Electrical and Electronics Engineers Inc.*, Mar. 2020, pp. 152–156. doi: 10.1109/ICIM49319.2020.244689.

[28] V. Uma Rani, M. Sreenivasa Rao, and M. C. Tech Scholar in, "Detection and Privacy Preservation of Sensitive Attributes Using Hybrid Approach for Privacy Preserving Record Linkage Kotra Sai Srujana," *International Journal on Recent and Innovation Trends in Computing and Communication*, 2017, [Online]. Available: http://www.ijritcc.org

[29] M. Franke, V. Christen, P. Christen, F. Rohde, and E. Rahm, "(Privately) Estimating Linkage Quality for Record Linkage," *in Advances in Database Technology - EDBT, OpenProceedings.org*, Nov. 2023, pp. 294–306. doi: 10.48786/edbt.2024.26.

[30] S. Thomas and J. Sluss, "Fraud detection through data sharing using privacy-preserving record linkage, digital signature (EdDSA), and the MinHash technique: Detect fraud using privacy preserving record links," *The Journal of Engineering*, vol. 2023, no. 12, Dec. 2023, doi: 10.1049/tje2.12341.

[31] D. Vatsalan, Z. Sehili, P. Christen, and E. Rahm, "Privacy-Preserving Record Linkage for Big Data: Current Approaches and Research Challenges." [Online]. Available: http://ec.europa.eu/justice/dataprotection/index

[32] Y. Lindell and B. Pinkas, "Secure Multiparty Computation for Privacy-Preserving Data Mining," 2008.

[33] "2019 View of Secure Multi-Party Computation in Genomics_ Protecting Privacy While Enabling Research Collaboration".

[34] M. Ostermann, I. Nesterow, and M. Wolfien, "A Hybrid-Approach for Privacy Preserving Record Linkage - A Case Study from Germany," *Stud Health Technol Inform*, vol. 316, pp. 43–47, Aug. 2024, doi: 10.3233/SHTI240340.

Citation of this Article:

Patel Krupali, & Dr. Prashant Pittalia. (2026). Privacy-Preserving Record Linkage: A Survey of Key Concepts and Approaches. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(1), 66-73. Article DOI https://doi.org/10.47001/IRJIET/2026.101008

****

