LexisNexis® RISK SOLUTIONS | Government

WHITE PAPER

How Privacy-Preserving Record Linkage Can Transform the Future of Data Exchange

Health and Human Services (HHS) agencies are tasked with the complex challenge of understanding their diverse populations to enhance their health and well-being. However, the data they rely on is often fragmented and stored across multiple systems, leading to gaps in critical insights.

As these agencies strive to better serve their members and ensure resources provided are reflective of the diverse needs of their populations, the benefits of cross-agency data sharing become obvious. Unfortunately, when working with personally identifiable information (PII), complicated privacy regulations and the protection of sensitive data prevent the necessary exchange of critical datasets across departments and agencies.

In a perfect world, agencies could seamlessly connect the dots across various datasets without compromising individual privacy. Privacy-Preserving Record Linkage (PPRL) offers an approach that enhances data sharing without compromising privacy, allowing agencies to unlock more comprehensive and precise insights while ensuring absolute protection of sensitive information.

By enriching records in a way that keeps identities secure, PPRL can empower agencies to make data-driven decisions that enhance the quality of care, improve policymaking and ultimately deliver better outcomes for the communities they serve.

Common challenges

Government agencies can face multiple obstacles in achieving a complete and accurate view of their populations and the underlying individuals they serve. The most pressing issues include:

Inconsistent data quality: Agencies may have messy, incomplete and inaccurate data that needs to be cleaned, deduplicated and standardized, often requiring resource and time-heavy processes.

Data fragmentation: An individual's data can be scattered throughout fragmented systems, preventing agencies from developing a comprehensive view of the people they serve.

Challenges establishing cross-agency collaboration: When looking to share individual data between each other, agencies can struggle with complicated data sharing requirements and regulations.

Process implementation: Effective use of data received from other agencies requires the establishment of clear and distinct workflows, which may not align with existing systems.

Data security concerns: Protecting identifiable data within and between agencies is paramount and requires navigating complex security regulations and restrictions.

Desire for comprehensive insights: By augmenting their own data with external sources, agencies can improve service delivery, streamline operations and ensure the appropriate allocation of resources.

Ensuring compliance: In sharing and using data, agencies must comply with relevant regulations such as the Health Insurance Portability and Accountability Act (HIPAA), the Gramm-Leach-Bliley Act (GLBA) and the Driver's Privacy Protection Act (DPPA).

Current solutions

Agencies looking to elevate the quality of their data currently have two primary options:

1. **Match and enrich data by sharing PII:** This approach provides for precise identity matching and data enhancing, enabling a more detailed understanding of an individual with a robust array of potential data sources. In situations where inaccurate data is present, there is a clear resolution process. When sharing PII, data can be used for functions such as research, enrollment or care management.

   However, matching and enriching an agency's population while sharing PII comes with significant privacy and security risk. Ensuring compliance and oversight can be onerous, in addition to adherence to multiple security and regulatory frameworks. New interagency memorandums of agreement/understanding (MOAs/MOUs) would need to be developed with each new party, and even for each new use case or project, requiring significant resource dedication across the agency. In addition, data governance and use case restrictions may limit the ability to freely share information across key parties, often watering down the insights that can be shared across agencies.

2. **De-identify data to preserve privacy:** De-identification minimizes privacy concerns and simplifies data sharing, but comes at the expense of data precision, leading to high error rates and poor matching quality. For research-based use cases, this lower-risk approach can be more suitable for analysis over time, particularly as it relates to policy and program analysis.

   While de-identified data may be helpful for historical analysis and be useful for research, it can fall short in delivering real-time, actionable insights linked and updated at the individual level.

The challenge is clear: agencies either compromise data quality for privacy or sacrifice privacy for actionable, individual-centric insights. Fortunately, with PPRL, we can help to reconcile these competing priorities.

Balancing privacy with actionable insights

In an ideal world, agencies would be able to share PII without fear of jeopardizing individual security and privacy. Agencies could capture and share insights amongst one another to offer a holistic view of individuals, ensuring the people they serve are getting the full scope of services they are entitled to, while protecting against misuse of government resources.

As new information on a person becomes available, it could be seamlessly integrated to more proactively and positively impact an individual's journey and adjust agency policies and processes based on data-driven insights. Unfortunately, this is not today's reality.

Instead, agencies often struggle to maintain multiple processes while sharing data. On the one hand, they receive fragments of individual information through identified workflows. On the other hand, they receive population-level, aggregate insights through completely different processes. In the end, they must juggle identified and de-identified data sharing workflows and underlying processes that do not fully accomplish their potential.

What is PPRL and how can it help?

PPRL offers a solution that balances the need for linking detailed, accurate data while maintaining and protecting individual privacy. The PPRL process takes the PII of an identity and replaces it with a non-sensitive, secure placeholder token that is a randomized string of characters. This token can then help to enable data matching and aggregation across datasets.

De-identification can help to minimize privacy issues and prevents the collection and storage of PII. This process also solves common identity problems such as confusing a father and son who share a first and last name or two individuals who may have overlapping identifiers. As each individual identity is tokenized, end users have the ability to keep a directional line of sight that is connected to the right person, even if some of the data has previously been commingled.

PPRL can safely transform data sharing, back and forth through future iterations, whether between agencies or between an agency and a private company.

For Health and Human Services agencies, using PPRL to share information can be powerful in a variety of use cases such as disease surveillance, health economics and outcomes research (HEOR), fraud detection, and program evaluation. PPRL can facilitate a comprehensive, real-time view of an individual's interactions with multiple agencies over time while maintaining strict privacy safeguards.

The benefits of PPRL

PPRL provides an opportunity for government agencies to safely exchange information and align benefits across programs, ensuring individuals receive the full spectrum of services they are entitled to, all while protecting their privacy.

The advantages of adopting PPRL are numerous:

Anonymized data sharing: By anonymizing beneficiary data at the source, PPRL can significantly reduce the risk of inappropriate data exposure.

Simplified privacy compliance: PPRL's process of removing PII may eliminate the need for complicated data use agreements with data exchange partners, easing the burden of compliance.

Broadened data sharing: PPRL can facilitate the sharing of data across agencies, jurisdictions and federal programs, ensuring the right resources get to the right people who need assistance.

Accurate deduplication: PPRL's precision in matching records can help identify duplicate records, improving data quality. In scenarios that involve case counts, such as immunization or disease surveillance, duplication can distort funding and resource allocation.

Enhanced record completion: By filling in gaps in datasets, PPRL can provide a more complete understanding of a person's journey and interactions with various agencies and programs.

Return on investment (ROI): Although implementing the infrastructure for PPRL requires an initial investment, it can be leveraged across multiple datasets and projects, providing long-term benefits and a strong ROI.

As public health challenges grow more complex and budgets tighten, agencies are under increasing pressure to deliver more sophisticated, efficient services without compromising privacy. By reducing duplicate records, improving data quality and streamlining cross-agency collaboration, PPRL helps agencies cut costs associated with manual reconciliation, redundant services and inefficient resource allocation.

PPRL empowers agencies to continue to enable data sharing while safeguarding the trust of the communities they serve.

The evolution of PPRL

PPRL is evolving from basic, cohort-level analysis – often used in research settings – to leveraging more sophisticated and effective referential linking.

Legacy PPRL – A traditional PPRL approach works in situations where there is no plan to share identified information and the use of the data is limited to identifying aggregate trends and analysis. While this method protects patient data, it can create large gaps in understanding since data is hard to match across identity variations and can be difficult to update in real-time.

These data gaps are lost opportunities for developing policy more proactively that is reflective of underlying individuals in a program population.

Current state PPRL has several additional shortcomings. Data cannot be accurately matched across identity variations and therefore, cannot be effectively enriched. In addition, the data is often stale and difficult to incorporate with new data sources as they become relevant. As more data is added to current PPRL processes, the matching rate can worsen, as users can only rely on the underlying PII available in each dataset. This can result in a high rate of false positives and false negatives, along with increasing frequency of missed linkages due to tokens that are reliant on point-in-time knowledge.

Future State PPRL – A future foundation of PPRL lies in the power of referential linking. This approach leverages a tokenization process – linking data via tokens – that factors in the variety of ways an individual has changed and will continue to change over time. A robust, referential dataset is a critical component of effective PPRL.

When leveraging referential data with PPRL, the data on an individual is precise, current and actionable, while fully de-identified for other parties. Referential data improves the accuracy of linking and matching, ensuring that linked records belong to the same individual. The use of referential data also allows for more robust data enrichment that can provide a comprehensive perspective of an individual. This can provide clarity in analysis for healthcare research and program evaluation.

For example, with referential linking driving more effective PPRL, agencies could be able to establish data sharing workflows that more effectively identify when an individual may already be on another benefit program's roster. This can lead to enhanced care management and care coordination, enhancing synergies and reducing costs between agencies for outreach, care and communication. With PPRL, only tokens are shared back and forth between agencies, providing for a seamless, secure flow of information that can enhance understanding and improves service delivery.

LexisNexis® Risk Solutions is a leader in referential data, with billions of public records, encompassing nearly half a century of identity variations.

Best practices for implementing PPRL

To maximize the benefits of PPRL, agencies should resolve point-in-time tokens to a referential token that includes an individual's identity variations over time, providing a reliable basis for matching. Agencies can then improve internal data management processes, such as deduplication and aggregation across data silos, while safeguarding sensitive PII.

These referential tokens would also serve as a "token of truth" for incorporating additional datasets in the future, serving as the bridge for data that is separated across agencies without sharing the details of PII.

With PPRL, each agency retains their PII, and their understanding of that identity based on the PII they hold. But via tokens, they can understand where those identities reside in other datasets, where they can have an interagency conversation, or where they can benefit from additional information available within the industry. Concurrently, agencies can continue to enhance a separate but aligned workflow that can accommodate decisioning needs with greater precision by leveraging identified data. One does not have to come at the expense of the other.

With PPRL, agencies can keep protected PII onsite while benefiting from enriched, de-identified insights. This is only possible with referential tokenization and a robust underlying referential dataset that drives the matching process to improve accuracy, without compromising privacy.

Choosing the right data management partner for PPRL

As PPRL continues to gain traction, HHS agencies should work with a data management partner that has a proven track record of collecting and managing referential data.

LexisNexis® Risk Solutions is a leader in referential data, with billions of public records, encompassing nearly half a century of identity variations. Our proprietary linking technology, LexID®, can create a comprehensive view of an individual with over 99.99% precision. LexisNexis Risk Solutions is uniquely positioned to solve the shortcomings of current state PPRL solutions through a novel approach to data linking and de-identification. The LexisNexis® Gravitas® token is a person-centric token that leverages our referential data layer to enable precision and confidence linking across datasets, while maintaining individual privacy.

To unlock the full potential of Privacy-Preserving Record Linkage and create a future where privacy and data quality coexist seamlessly, LexisNexis Risk Solutions can help.

For more information, call 800.869.0751 or visit risk.lexisnexis.com/government/medicaid-program-management

Scan for Information

LexisNexis® RISK SOLUTIONS | Government

About LexisNexis® Risk Solutions

LexisNexis® Risk Solutions harnesses the power of data, sophisticated analytics platforms and technology solutions to provide insights that help businesses across multiple industries and governmental entities reduce risk and improve decisions to benefit people around the globe. Headquartered in metro Atlanta, Georgia, we have offices throughout the world and are part of RELX (LSE: REL/NYSE: RELX), a global provider of information-based analytics and decision tools for professional and business customers. For more information, please visit www.risk.lexisnexis.com and www.relx.com.