Selfii.

White paper: Enabling Medical Research Using Health Information Exchange Data

Table of Contents:

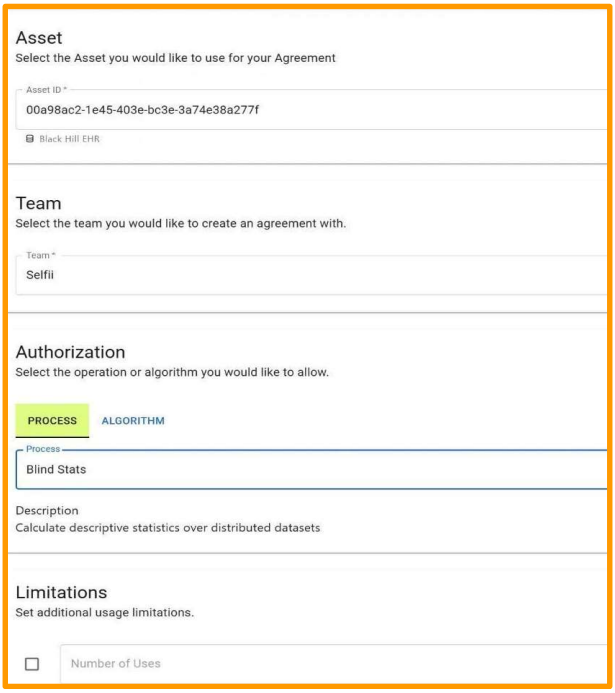
Executive Summary

Medical research depends on high-quality, real-world clinical data, yet Health Information Exchanges (HIEs) struggle to enable access to this data securely and efficiently. Traditional methods of de-identification and data transfers strip away critical insights, introduce compliance risks, and make data sharing inefficient and expensive. HIEs, as aggregators of rich clinical data, are uniquely positioned to support important research—but they need a solution that enables access without compromising privacy, security, or compliance.

Many HIEs today rely on de-identification, tokenization, and data transfers to enable research collaborations. While these methods are designed to protect patient privacy, they often result in significant data loss, operational inefficiencies, compliance risks and additional expense. De-identification techniques strip away valuable clinical insights, particularly in free-text physician notes, DICOM imaging files, and genomic data, rendering them far less useful for advanced analytics and AI-driven models. Additionally, once data leaves the HIE environment, control is lost, increasing the risk of unauthorized access, re-identification, and compliance violations under HIPAA, CCPA, and GDPR.

Selfii Exchange: A Secure, Full-Fidelity Alternative

Selfii Exchange, powered by TripleBlind's privacy-preserving technology, introduces a transformative approach to research data access. Instead of requiring de-identification or physical data transfers, Selfii Exchange enables secure, in-place access to high-fidelity data, allowing researchers to derive insights without ever viewing raw patient records.



Graphic 1: A snippet of the Agreements section of the Exchange app. HIEs have full configurability on which data assets are usable (each data set gets its own asset ID). The HIE can similarly configure who has access to any asset, what operations are allowed on an asset, how many times an operation is allowed to be run, and what if any fees are associated with performing the operation.

With Selfii Exchange, HIEs maintain full control over their data while enabling real-time, privacy-preserving research collaboration. Researchers can query structured and unstructured datasets—including pathology reports, cognitive assessments, and prescription histories—without exposing PHI. Unlike traditional methods, Selfii Exchange supports multiparty computation (MPC) to ensure that data remains encrypted during computation, allowing researchers to extract key insights while keeping patient information secure.

Key Benefits for HIEs

- Eliminate De-Identification. HIEs can avoid the costly expenses of de-identification, tokenization, and expert determinations of HIPAA compliance.
- Preserve Data Fidelity. Researchers can analyze full clinical datasets, including complex PDFs, physician notes, and imaging metadata, without compromising patient privacy.
- Eliminate Data Transfers. Data never leaves the HIE environment, reducing the risks associated with breaches, re-identification, and compliance violations.
- Ensure Compliance. Meets HIPAA Expert Determination standards and aligns with CCPA, GDPR, and other privacy regulations.
- Maintain Full Control. HIEs define who can access data, what they can see, and how they can use it with granular agreement controls and audit logs.
- Enable Faster Research. Researchers can immediately access structured and unstructured clinical data without waiting for de-identification or manual approvals.

With Selfii Exchange, HIEs can bridge the gap between research and real-world clinical data, enabling groundbreaking discoveries while ensuring data privacy, security, and compliance.

The Researcher Data-Sharing Challenge for HIEs

Growing Demand from Researchers

Research institutions increasingly depend on real-world clinical data to drive medical advancements, evaluate treatment efficacy, and conduct population health research. Access to high-quality, diverse patient data enables researchers to identify trends, develop precision medicine strategies, and refine predictive models that improve patient outcomes. As central aggregators of health information, HIEs play a critical role in facilitating these research efforts. By consolidating data from hospitals, clinics, laboratories, and other healthcare entities, HIEs serve as a valuable bridge between data providers and the research community, offering a comprehensive view of patient health that is essential for robust, large-scale studies.

Recognizing this need, many HIEs are actively collaborating with researchers or exploring partnerships to make their data available for scientific and clinical studies. Researchers urgently need access to diverse clinical datasets, including structured electronic health records, unstructured physician notes, and medical imaging files such as DICOM scans and pathology reports. However, despite this growing demand, existing data-sharing models create significant barriers that make it difficult for HIEs to provide researchers with high-quality, privacy-compliant data. Legacy approaches often require data de-identification, tokenization, or physical data transfer, each of which introduces compliance risks, operational inefficiencies, and loss of clinical fidelity. As a result, HIEs struggle to balance the need for secure data access with the obligation to protect patient privacy and meet regulatory requirements—a challenge that traditional solutions have yet to fully address.

Why Traditional Approaches Fall Short

Traditional research data-sharing solutions, including those offered by Datavant and other tokenization-based providers, rely on de-identification or anonymization to comply with privacy regulations before data can be shared with researchers. While these methods are designed to remove personally identifiable information (PII) and protect patient privacy, they often come at a cost. One of the greatest challenges with traditional de-identification and tokenization methods is that HIEs lose control over their data once it is shared. Even after removing direct identifiers, there remains a residual risk of re-identification, particularly for datasets involving rare diseases, small patient populations, or longitudinal studies where records can be linked across multiple sources. HIPAA's Safe Harbor method, which mandates the removal of 18 specific identifiers, does not fully eliminate these risks. Advanced re-identification techniques—such as cross-referencing de-identified data with publicly available information—have been shown to successfully reconstruct patient identities in certain cases. This risk is further exacerbated when data is transferred outside the HIE's controlled environment and stored in external research repositories, where its use and further dissemination become difficult to monitor or restrict.

Beyond security and compliance risks, traditional de-identification methods significantly reduce data quality, making datasets less valuable for researchers. Removing personally identifiable information (PII) often requires eliminating or altering free-text physician notes, scanned documents, and metadata from medical imaging, stripping away clinically relevant details that researchers rely on for analysis. Medical images, such as DICOM scans and pathology slides, pose a unique challenge because they often contain embedded patient identifiers and cannot be meaningfully de-identified without disrupting the integrity of the data. Similarly, handwritten physician notes and PDF reports are difficult to process using automated de-identification techniques, leading to either excessive redaction or outright exclusion from datasets. The result is a compromised dataset that lacks the depth necessary for meaningful AI modeling, predictive analytics, and real-world evidence studies.

Even structured electronic health record (EHR) data—when tokenized or de-identified—loses its full research potential. Tokenization attempts to maintain record linkages but often adds complexity and fragmentation, making large-scale research difficult. Without key demographic and contextual details, researchers struggle to assess disease progression, treatment efficacy, and social determinants of health. By relying on these traditional approaches, HIEs are forced to choose between privacy and data usability, often leading to missed opportunities for valuable research partnerships. Without a better solution, they risk losing control of their data, diminishing its value, and facing ongoing compliance uncertainties.

Regulatory Considerations

Navigating the complex landscape of healthcare data privacy regulations presents a significant challenge for HIEs seeking to support research collaborations. Under HIPAA, de-identification is typically achieved through one of two approaches: Safe Harbor or Expert Determination. The Safe Harbor method requires the removal of 18 specific identifiers, often leading to a loss of clinical detail that limits the utility of datasets for research. Alternatively, the Expert Determination method allows a qualified professional to certify that the risk of re-identification is "very small," but this approach requires ongoing risk assessments, making it time-consuming, costly, and difficult to scale. Even when these methods are applied, data still must be physically transferred to researchers, creating additional security vulnerabilities.

The introduction of state-level and global privacy laws has further complicated how HIEs manage data access for research. Regulations such as the California Consumer Privacy Act (CCPA) and the General Data Protection Regulation (GDPR) impose strict requirements on the handling of personal health data, including limitations on sharing—even when de-identified. Many of these laws define de-identified or pseudonymized data as personal data, meaning HIEs must implement additional safeguards to prevent improper use. This has made it increasingly difficult for HIEs to balance supporting medical research with maintaining full regulatory compliance.

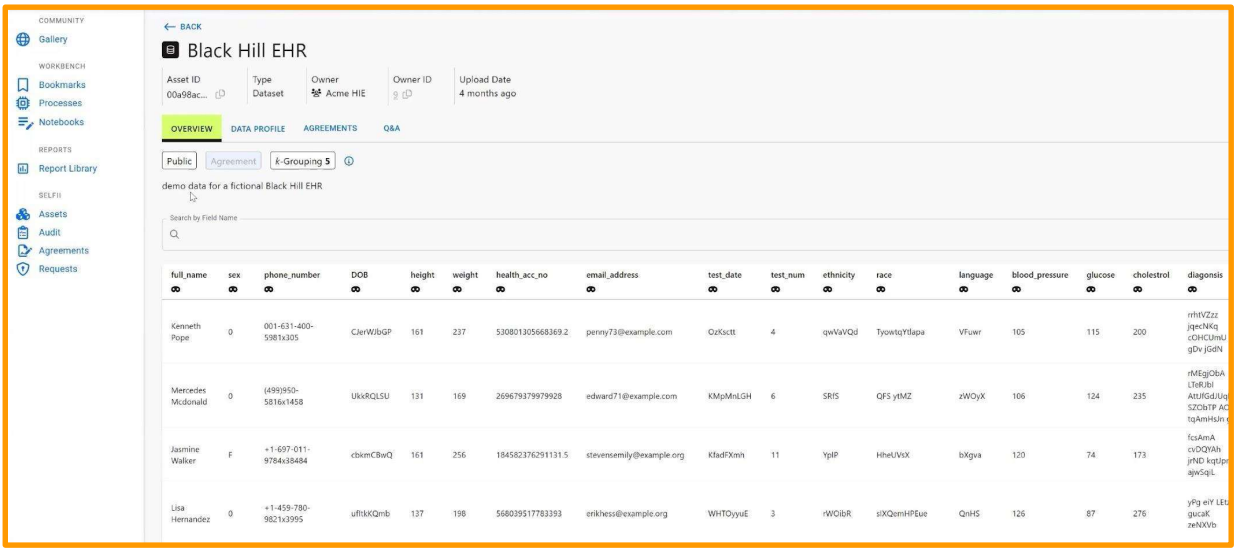
Selfii Exchange, powered by TripleBlind's secure multiparty computation (MPC), eliminates these compliance risks by ensuring that raw patient data never leaves the HIE environment. Instead of relying on de-identification or data transfer, researchers can perform analyses within a privacy-preserving framework, where access is tightly controlled. Additionally, detailed audit logs track every data request and researcher interaction, ensuring full transparency and accountability in compliance with HIPAA, CCPA, GDPR, and other evolving data protection regulations. This privacy-first approach allows HIEs to support research while maintaining control over their data, ensuring compliance without compromising data integrity.

Selfii Exchange: A Secure, Full-Fidelity Alternative

How It Works:

Selfii Exchange with TripleBlind's privacy-enhancing technology transforms how HIEs facilitate research by ensuring that data remains in place, eliminating the risks associated with data transfer, duplication, or external storage. Instead of extracting and de-identifying records before sharing them, researchers interact with the data securely within the HIE's controlled environment, preserving its original structure, quality, and clinical integrity. This approach prevents data breaches, re-identification risks, and compliance violations that often arise when datasets are moved between organizations.

Unlike traditional de-identification methods that remove or alter key clinical details, Selfii Exchange enables researchers to access full-fidelity data without ever seeing raw patient information. Whether working with structured EHR data, free-text physician notes, PDFs, or medical imaging files, researchers can perform advanced analyses while the system automatically enforces privacy controls. Sensitive elements are never exposed, ensuring compliance while still allowing meaningful insights to be extracted from even the most complex data sources.



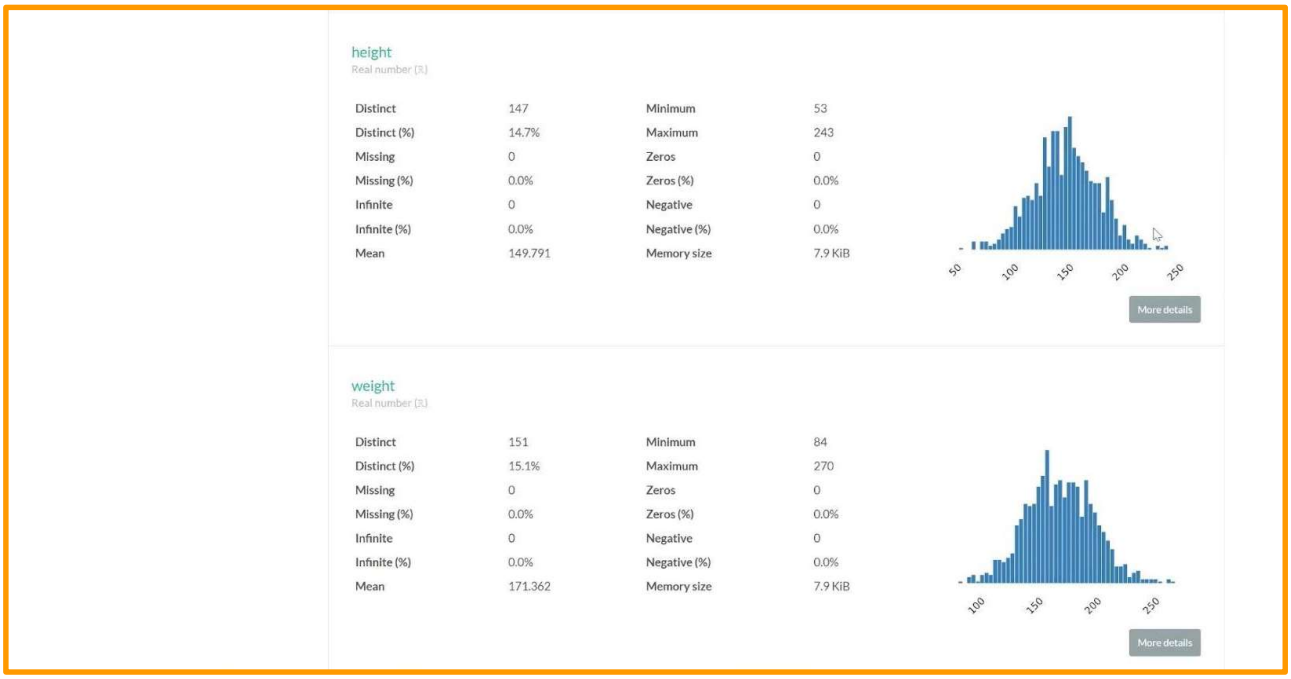
Graphic 2: Researchers can browse masked overviews of the available datasets to pre-determine if one suits their needs. The masked categories show automatically generated synthetic examples of each column of data in Exchange.

Each research query is processed securely, returning only the permitted outputs based on predefined agreements set by the HIE. This ensures that researchers receive the information necessary for their studies without the ability to reconstruct or re-identify individual patient records. By maintaining data integrity, eliminating movement risks, and enforcing strict access

controls, Selfii Exchange enables a new model of research collaboration—one that prioritizes both data privacy and scientific progress.

Key Capabilities for HIEs

Selfii Exchange offers advanced capabilities that empower HIEs to securely collaborate with researchers while maintaining full control over their data. One of its most powerful features is its Refinery for Complex Data, which enables the system to process unstructured and difficult-to-de-identify information such as medical images, PDFs, and handwritten physician notes. Instead of stripping away valuable clinical details, the refinery extracts only the necessary structured insights—such as clinical codes, diagnoses, and relevant biomarkers—while ensuring Protected Health Information (PHI) remains secure. This allows researchers to gain meaningful insights from complex datasets without direct access to raw patient records.



Graphic 3: The "quick statistics" feature in which users can see the distributions within the data to get a better idea of the set they are considering at a glance.

HIEs retain complete authority over data access through granular agreement controls, defining exactly who can access data, for what purpose, and under what conditions. Unlike traditional data-sharing agreements that involve static de-identified datasets, Selfii Exchange enables dynamic, real-time access while enforcing strict permissions, auditability, and usage restrictions. This ensures that each research query adheres to regulatory and contractual obligations, protecting both the HIE and patient privacy.

Scalability is another key advantage of Selfii Exchange, as it is designed to handle diverse datasets across multiple sources and locations. Whether working with large-scale genomic data, longitudinal patient records, or regional health registries, the platform scales to meet growing data demands without compromising security or computational performance. Beyond security and compliance, Selfii Exchange also presents new revenue opportunities for HIEs, allowing them to monetize high-quality datasets responsibly by enabling AI developers, pharmaceutical researchers, and academic institutions to securely analyze data without ever taking possession of it. By integrating privacy-enhancing technologies such as secure multi-party computation and blind learning, Selfii Exchange not only safeguards patient data but also accelerates medical innovation, paving the way for the next generation of privacy-preserving healthcare research.

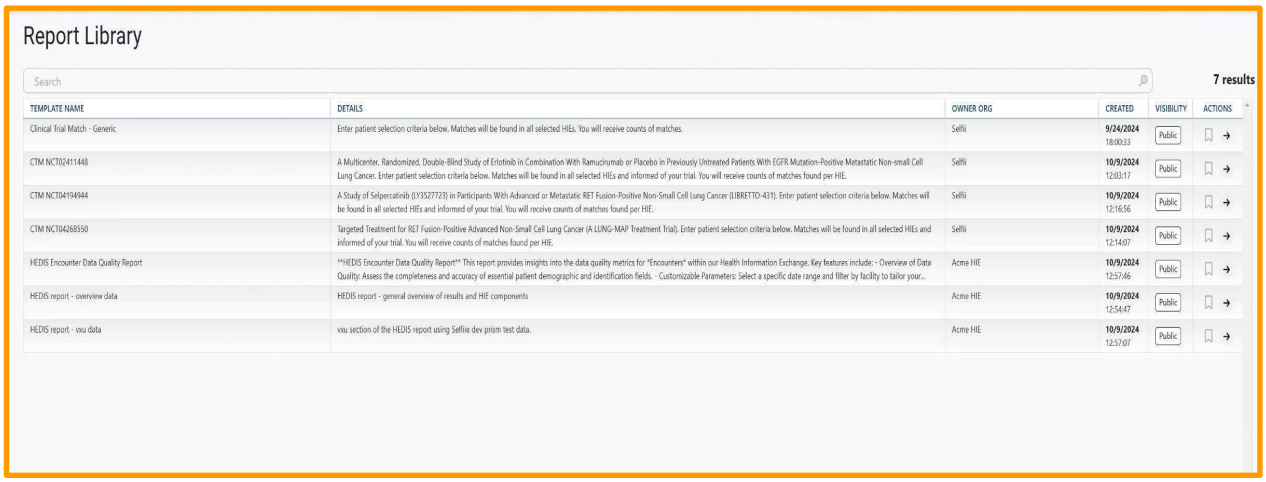
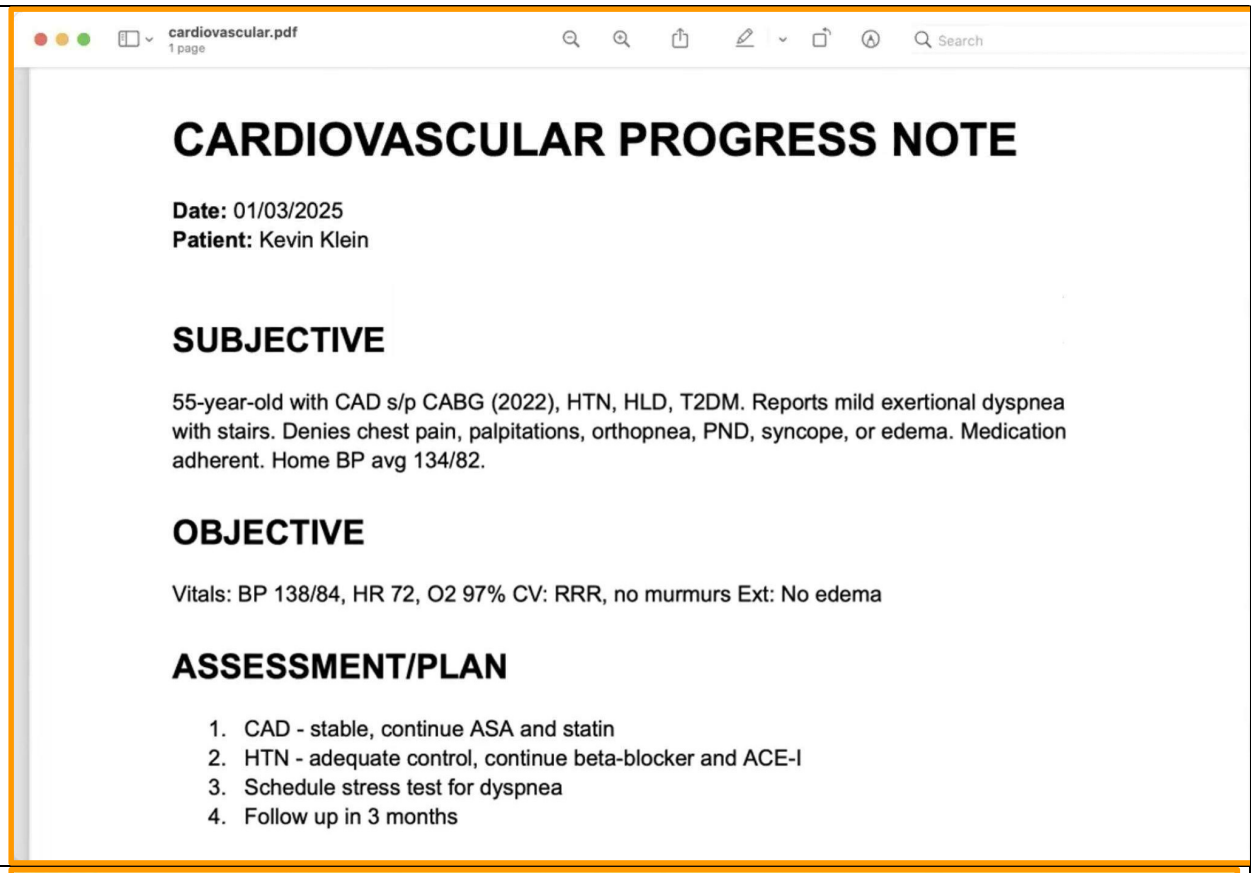


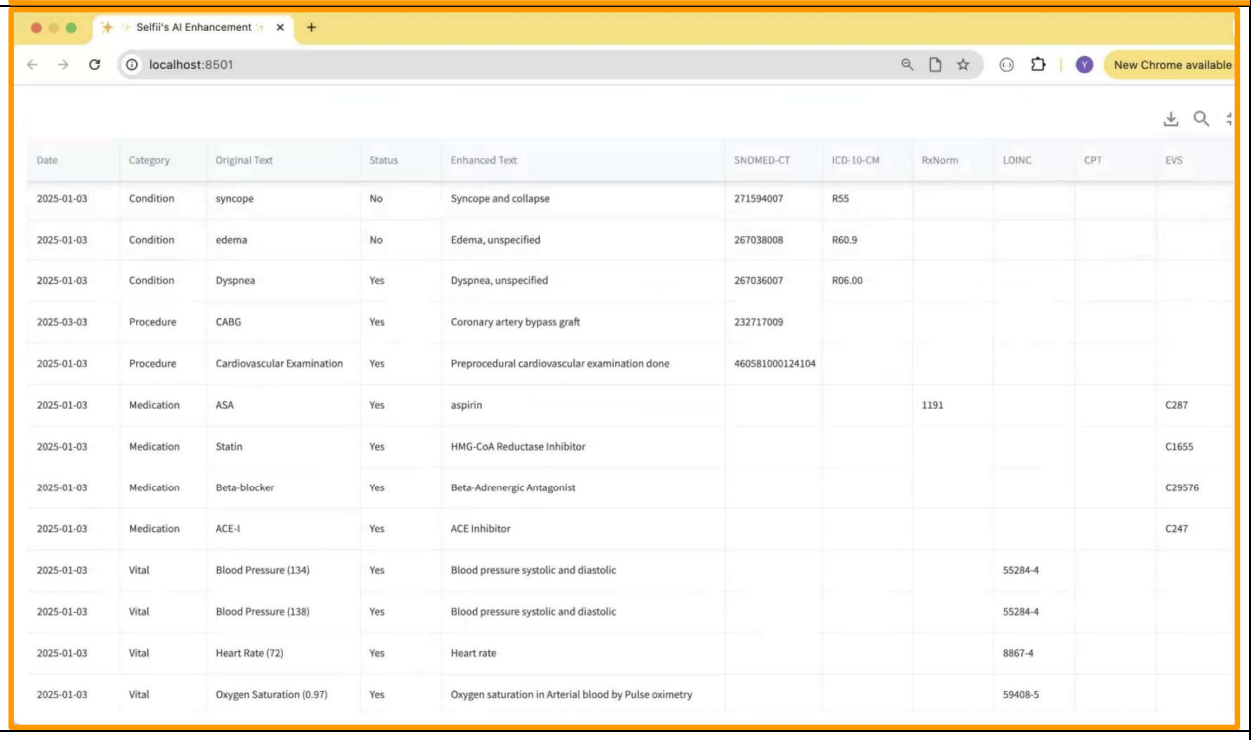
Figure 4: The report library tracks dates and times of reports run on datasets, and by whom. HIEs can allow researchers to build and run their own reports, knowing the PHI is protected, and also knowing how frequently it has been used.

For example, some pharmaceutical research requires data with geographic diversity. These studies therefore require data from two or more HIEs, where any given HIE is only supplying a portion of the returned information. The solution easily enables this scenario, and it also makes each HIE's dataset even more valuable because the researchers can work as if accessing a single data set when in fact it might be many distinct HIE sources on Selfii Exchange.

Privacy and Compliance Benefits for HIEs

Selfii Exchange with TripleBlind's privacy-enhancing technology offers a regulatory-backed security framework that ensures HIEs can support research collaborations without compromising patient privacy or compliance. Unlike traditional de-identification methods, which rely on removing or altering patient data, Selfii Exchange leverages a mathematically proven approach that allows researchers to perform computations on encrypted data without ever accessing or seeing the raw records. This groundbreaking technology has been independently validated by privacy expert Dr. Bradley Malin, confirming that it meets HIPAA's Expert Determination standard and aligns with evolving global privacy laws such as CCPA and GDPR.





Date	Category	Original Text	Status	Enhanced Text	SNOMED-CT	ICD-10-CM	RxNorm	LOINC	CPT	EVS
2025-01-03	Condition	syncope	No	Syncope and collapse	271594007	R55				
2025-01-03	Condition	edema	No	Edema, unspecified	267038008	R60.9				
2025-01-03	Condition	Dyspnea	Yes	Dyspnea, unspecified	267036007	R06.00				
2025-03-03	Procedure	CABG	Yes	Coronary artery bypass graft	232717009					
2025-01-03	Procedure	Cardiovascular Examination	Yes	Preprocedural cardiovascular examination done	460581000124104					
2025-01-03	Medication	ASA	Yes	aspirin			1191			C287
2025-01-03	Medication	Statin	Yes	HMG-CoA Reductase Inhibitor						C1655
2025-01-03	Medication	Beta-blocker	Yes	Beta-Adrenergic Antagonist						C29576
2025-01-03	Medication	ACE-I	Yes	ACE Inhibitor						C247
2025-01-03	Vital	Blood Pressure (134)	Yes	Blood pressure systolic and diastolic				55284-4		
2025-01-03	Vital	Blood Pressure (138)	Yes	Blood pressure systolic and diastolic				55284-4		
2025-01-03	Vital	Heart Rate (72)	Yes	Heart rate				8867-4		
2025-01-03	Vital	Oxygen Saturation (0.97)	Yes	Oxygen saturation in Arterial blood by Pulse oximetry				59408-5		

Figure 5: Our "Refinery" AI can ingest and understand PDFs and other text documents, and output into structured data tagged with proper medical coding. In the PDF above the signs, symptoms, medications, etc. are found and codified.

By eliminating the need for hardware-dependent secure enclaves or synthetic datasets, Selfii Exchange provides a more scalable, cost-effective, and flexible solution for data access. Many traditional privacy-preserving methods require specialized infrastructure, making them expensive and difficult to implement across multiple organizations. In contrast, Selfii Exchange operates entirely within the HIE's existing environment, ensuring that data remains in place while allowing approved researchers to derive insights from real-world clinical data. This means that HIEs can provide full-fidelity data access without assuming the risks associated with physical data transfer, decryption, or re-identification threats.

To maintain transparency and accountability, every researcher interaction is fully logged and auditable, giving HIEs a complete record of how, when, and why their data is being accessed. These detailed audit trails enhance security, support compliance reporting, and provide clear oversight of all research activities. With Selfii Exchange, HIEs no longer must choose between enabling research and protecting patient privacy—they can confidently support innovation while maintaining complete regulatory compliance and control over their data.

Real World Uses Cases: How HIEs Can Immediately Support Ongoing Research

HIEs hold a wealth of clinical data that researchers urgently need to advance studies in oncology, neurology, cardiology, and more. However, traditional methods of de-identification, data movement, and restrictive access policies make it difficult for researchers to obtain high-fidelity, privacy-compliant data in real-time. With Selfii Exchange, researchers can immediately access structured and unstructured HIE data while ensuring that patient privacy remains fully protected.

Use Case 1: Oncology Researchers Seeking Immediate Access to Pathology Reports

Cancer researchers are currently working to develop better tumor staging models and precision medicine strategies, but many struggle to access high-quality pathology reports due to de-identification barriers. Traditional approaches require removing free-text clinical descriptions, metadata from imaging files, and molecular biomarker details, resulting in fragmented datasets that lack critical insights.

With Selfii Exchange, researchers could immediately query pathology reports, extract tumor size, histological grade, and biomarker status while ensuring that raw images and PHI remain protected. This would enable faster development of AI-driven models for tumor detection, treatment response, and patient survival predictions—accelerating cancer research while keeping HIE data secure and under full regulatory compliance.

Use Case 2: Neurology Researchers Needing HIE Data for Alzheimer's Studies

A major challenge in Alzheimer's research is identifying early warning signs that could help predict cognitive decline before symptoms become severe. Researchers studying the progression of neurodegenerative diseases need access to longitudinal clinical notes from neurologists, cognitive assessment scores, and MRI brain scans to develop more effective early detection models. However, due to data access restrictions and de-identification risks, many studies face long delays in obtaining the necessary clinical datasets.

With Selfii Exchange, researchers could immediately access structured insights from physician notes and MRI metadata—allowing them to analyze cognitive function trends, brain structure changes, and early biomarkers of Alzheimer's. This privacy-preserving access would eliminate the need for manual data transfers and de-identification, allowing real-time AI and statistical modeling to drive early intervention strategies that could slow disease progression and improve patient outcomes.

Use Case 3: CV Researchers Needing Rapid Access to Prescription and Lab Data

Cardiovascular disease remains the leading cause of death worldwide, and researchers are racing to determine which medications are most effective in preventing heart attacks and strokes in high-risk patients. To conduct meaningful studies, scientists require access to prescription adherence records, lab results, and demographic data to assess how different populations respond to treatment over time. However, de-identification often strips out key demographic and clinical details, making it difficult to perform precision medicine research that accounts for genetic, ethnic, and socioeconomic factors.

With Selfii Exchange, researchers could immediately analyze longitudinal prescription data, cholesterol levels, and patient outcomes across diverse patient populations, enabling them to identify the most effective treatment regimens while maintaining full compliance with HIPAA, CCPA, and GDPR. This real-time access to secure, high-fidelity data would support the development of personalized treatment strategies, helping reduce the incidence of heart attacks and improve patient survival rates.

Bridging the Gap Between HIEs and Research Innovation

Researchers across multiple disciplines urgently need access to rich, real-world clinical datasets that HIEs already possess. However, current data-sharing barriers prevent them from obtaining high-fidelity data in a timely and secure manner. With Selfii Exchange, researchers could immediately access structured and unstructured clinical data within the HIE environment, eliminating delays and compliance hurdles while ensuring that patient privacy remains protected.

By enabling real-time, privacy-preserving data access, Selfii Exchange allows HIEs to support medical breakthroughs in oncology, neurology, cardiology, and beyond—without sacrificing security, compliance, or data integrity.

Get Started with Selfii Exchange

HIEs have a unique opportunity to support groundbreaking research while ensuring data privacy, security, and regulatory compliance. By adopting Selfii Exchange with TripleBlind's privacy-preserving technology, HIEs can immediately enable secure researcher access to high-fidelity clinical data without the risks associated with de-identification, data transfers, or loss of control.

Whether your organization is already collaborating with researchers or exploring new ways to monetize data responsibly, Selfii Exchange provides the technology to make it possible—securely, efficiently, and in full compliance with HIPAA, CCPA, and GDPR. Our platform is designed to eliminate barriers to research, allowing HIEs to drive medical innovation while maintaining full oversight of how data is accessed and used.

Let us transform research together - request a demo to see how Selfii Exchange enables real-time, privacy-preserving data collaboration.