

Empowering Your Data
Securely. Efficiently. Responsibly.



Table of Contents

🌀 Introduction	3
🌀 SapienSecure Enterprise Platform	4
🌀 Data Aggregation Technology	7
🌀 Medical Data De-identification	7
🌀 Natural Language Processing and Understanding	8
🌀 Modules	11
🌀 Case Studies	13
🌀 The Future	14

Current Healthcare Spending is **Unsustainable**

In the United States, based on Government estimates, **\$0.25 to \$0.33** of every dollar goes toward operations + administration at a cost of **\$1 Trillion annually** in healthcare. Recent statistics evaluate dirty data costs the U.S. healthcare industry **\$300 billion** or more and U.S. Attorney suggesting that **14%** of industry expense disappears through data mismanagement.

SapienSecure® enables real-time data access & discovery of population data across disparate databases that drive productivity and profitability. Our proprietary Natural Language Processing and Artificial Intelligence computer-driven models extract, classify, and index medical records for the purpose of accelerating access, and improving the quality, security, and integrity of data.

Furthermore, **less than 1% of health data** is used beyond direct patient care. Hospital data is locked in disparate databases across the health system and unavailable for external and internal data usage for administrative or clinical research. This includes critical R&D, diagnostics, and/or optimization of clinical systems, such as; billing, scheduling, and quality auditing efficiencies.



Primary Data



Secondary Data

Hospitals trust SapienSecure to safely access real-time medical data for analysis and audits to make data available as a de-identified export without having to remove or replace existing systems.

By building intelligent datasets physicians, researchers and administrative staff have better insight and predictability of patient outcomes and are empowered to leverage data securely across the healthcare system.

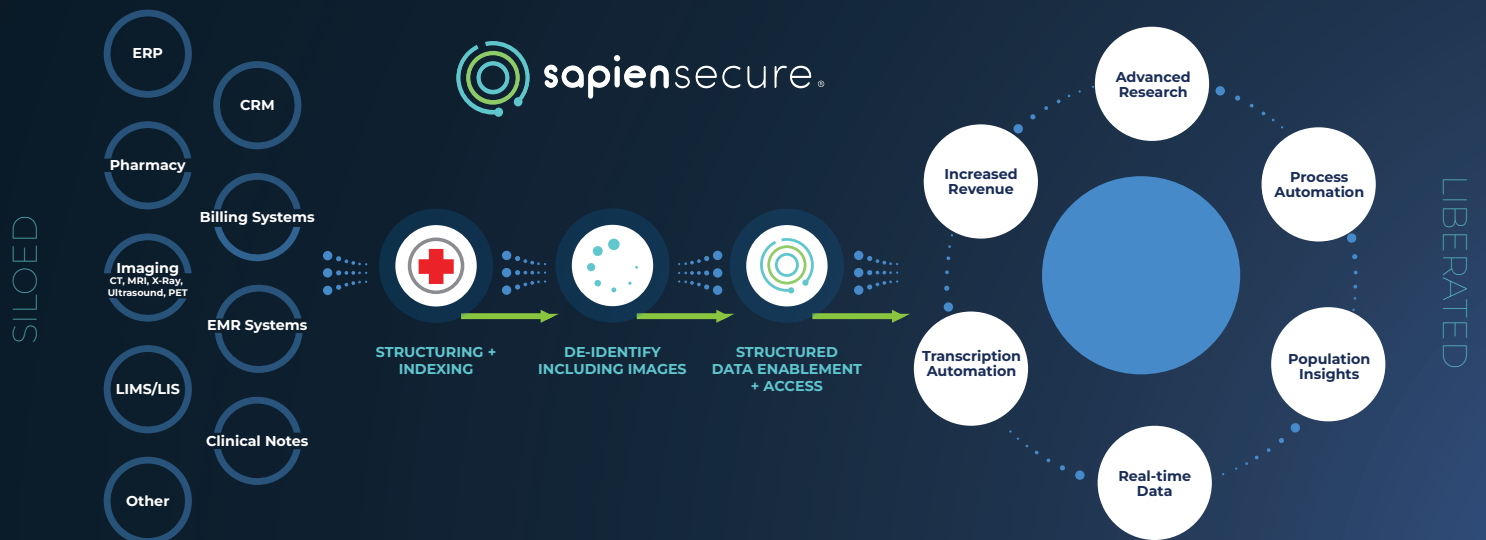
Securely. Efficiently. Responsibly.

SapienSecure Machine Learning technology utilizes NLP and AI to create intelligent datasets, compliant with patient privacy laws. This allows hospitals to clearly visualize, search, and analyze population-level data across multiple sources within the healthcare system to maximize efficiency and positively affect patient outcomes.

- 🎯 Improved access to en-masse population data for advancements in R&D
- 🎯 Optimize operations to increase hospital revenue and resource allocation
- 🎯 Delivering insights that ultimately improve patient outcomes and care

SapienSecure Enterprise Platform

SapienSecure is a simple, yet powerful, enterprise medical data management solution that empowers health systems to do more with their data.



Value in Volume

- Better Research Outcomes
- Bigger Breakthroughs
- Greater Monetization



Disparate Databases

SapienSecure is installed at source: inside medical facilities, server environments, local data centers, secure cloud configurations or in preparation to securely move to the cloud. Our enterprise software is installed and exists behind the firewall, where a health system already manages and stores medical data, and is entirely controlled by the health system's IT and Data Governance departments. Built on trust, our strong two-way collaboration is required to fully leverage the value and potential of SapienSecure. SapienSecure is configured, deployed and automated directly inside the health systems to provide full control of the software with complete administration of user access and permissions; not even SapienSecure team members can access the software without the hospital system's permission. This level of administration and service is why hospital systems trust SapienSecure with their most sensitive data.

SapienSecure's set up process is simple, and interoperable, taking only days to install and begin the steps to index, identify and classify hundreds of millions of patient records. The SapienSecure Enterprise Software is designed to plug into numerous disparate databases across a health system, spanning everything from electronic medical records and laboratory data to operations and finance. However large or segregated across the health system, SapienSecure is designed to provide a discoverable window into clinical data in a way that

can be easily digested by non-IT teams such as: executives, physicians, researchers, and analysts who ask pertinent questions, and seek real-time answers from billions of data points.

ELECTRONIC MEDICAL RECORDS:

EMR/EHR databases utilize a standard data transfer methodology; HL7 and FHIR. These databases hold comprehensive encounter data transcribed directly during point of care. Trickle down information or orders, such as diagnoses, coding systems, free-text reports, and notes are stored in these EMR/EHRs and made minimally accessible without sophisticated technology and SME. SapienNLP technology ingests millions of lines of unstructured text data per minute, and analyzes it for critical information that is then structured and indexed for dynamic query retrieval when using the SapienSecure dashboards.

MEDICAL IMAGING:

One of the largest databases in a hospital system relates to medical images, which stores petabytes of data at even a smaller hospital system. There is a vast amount of potential stored away in DICOM data, both from a radiologic and pathologic perspective. Much of this unstructured pathology and radiology diagnostic reports include key information (diagnoses, measurements, anatomy, procedures) and add tremendous value to datasets when extracted and structured.

CLINICAL NOTES:

Clinical interpretations are a critical piece of the medical chart, though are inherently difficult to extract due to the high volume of unstructured, and therefore, unsortable data that is contained within them. Subject matter expertise, and typically, manual data entry, is required to interpret and contextualize these unstructured notes in order to extract value from this highly specific medical language. Using SapienNLP technology, we can extract key elements from these notes and eliminate the need for manual data entry. Because this data becomes structured in the process, it is then technically possible to then associate it to studies, labwork, or patient visits and augment our interpretation of those elements.

BILLING SYSTEMS:

Billing in a vacuum is just a series of numbers. When attaching billing to the clinical study or imaging or the patients themselves, we can begin to create a foundation on which we can support audits and vastly improved billing practices. By being able to extract information from clinical notes or clinician reports relevant to billable procedures, billing procedures can be made more accurate and enable patient tracking through their care journey.

PHARMACY:

Pharmacy databases contain as much structured information as unstructured, as well as some of the most sensitive information about patients that could be available in a hospital. It is critical that the data that is extracted from any pharmacy database is efficient and de-identified, while also preserving as much information as possible.

LAB INFORMATION SYSTEMS:

Lab data is an incredible source of information but the sheer volume of data that is contained in any Lab Information System limits its use, especially when trying to examine data independently without inter-associations within the patient chart. By utilizing aggregation techniques, SapienSecure can map lab data to other information systems and break down the barriers that make this massive dataset such a daunting task.

ENTERPRISE RESOURCE PLANNING SYSTEMS:

ERP systems are critical infrastructure for ensuring daily operations and ensuring staff can focus on what they do best. By integrating operational databases with clinical databases, operations can be empowered with more information to make better decisions. More details, when appropriately sourced and coded, can only ever improve the decision-making power of management.

Data Aggregation Technology

By aggregating and indexing all of these database systems, your various business units can utilize SapienSecure Discovery as a clinical search engine, pulling relevant data from the network of associations to answer important clinical, operational, or research questions.



Indexing

Indexing is the process of scanning through billions of medical records in the background, gathering an understanding of where data is stored, how much data there is, and what types of data are available. Upon installation, SapienSecure begins the indexing process, automating the cataloging of data to enable quick search and discovery. When a user starts typing into the SapienSecure search field, the dynamic discovery tool leaps into action bringing forward relevant data, ready at your fingertips. This would not be possible without indexing. It's something that makes SapienSecure unique.



Structuring the Unstructured

SapienSecure extracts structured and unstructured medical data. Structured medical data includes tables of cleanly organized data, with aligned rows and columns. The fields are defined and easily sortable. Unstructured data includes medical images, free-language text reports, notes, voice recordings, videos, etc.; any data that isn't already organized and historically has required intelligent manpower to analyze and extract useful information from. SapienSecure minimizes the need for personnel to spend countless hours extracting data and organizing it, and instead simply requires a human-in-the-loop. SapienSecure indexes unstructured data, and does 80-90% of the heavy lifting, so that the hired personnel only need to confirm SapienSecure's findings, freeing up time for other critical tasks.

Medical Data De-Identification

SapienSecure was founded by two doctors and a medical engineer, whose foray into software development was a de-identification engine for medical data. It was critical in the early days to ensure that data being accessed for research and secondary use was safe and kept patient identities private.

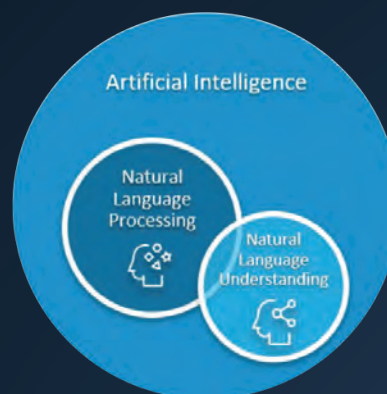
The SapienSecure De-ID system has come a long way since the early days as a small group of three, and is now touted by many to be industry leading. Its automated de-identification system supports vast types of medical data. It serves medical institutions, health data service providers and commercial health research companies.

Key features include:

- © Handwriting redaction: AI detection of handwriting and signatures for automatic quarantine.
- © Image Defacing: AI detection of faces on medical images, and automatic obfuscation by removal of the skin pixel data.
- © Medical record number shifting consistency: The detection and automatic tokenization of unique identifiers, dates, addresses, etc., that are then shifted to retain internal records consistency without retaining PHI.
- © Burnt-in Text Redaction: AI detection of text in medical images and OCR detection and redaction of PHI.
- © PHI NLP: Personal Health Information (PHI) redaction in free-written text (sentences, paragraphs, reports, etc.), using named entity recognition Natural Language Processing (NLP).
- © Differential Privacy and K-Anonymity: The right combination of indirect patient identifiers (age, sex, postal/ZIP code, ethnicity, blood type, etc.), can directly identify individuals; therefore SapienSecure aggregates tabular data using differential privacy algorithms.

- © Optical Character Recognition: Automated detection of text in forms for extraction and aggregation of contained data.
- © DICOM standard De-ID: Imaging meta-data de-ID.

The physician founders of SapienSecure were first authors on national standards white papers, including by the Canadian Association of Radiology AI working group de-ID white paper.



Natural Language Processing + Understanding

AI: NLP / NLU

What is Natural Language Understanding /Processing?

The aim of NLP is to process free-form natural language text so that it gets transformed into a standardized structure. Natural language understanding and/or processing is a key component of SapienSecure and what separates us from the rest of the pack. The ability for machines to understand is becoming a reality with the advent of artificial intelligence. SapienSecure uses dozens of NLU/NLP models to add structure to unstructured data.

There is a vast amount of information locked behind medical free-text, both from a historic perspective and ongoing using modern voice-to-text dictation into many different medical records systems (EMR, EHR, DICOM, etc.). From the computer's point of view, any natural language is free-form text. That means there are no set keywords or positions when providing an input.

For example, consider these three sentences:

- ③ What is the weather like tomorrow?
- ③ Will it snow in New York tomorrow?
- ③ Should I bring an umbrella with me tomorrow?

All of these sentences ask the same question, meant to understand what the weather will be like tomorrow. As homo sapiens, we identify underlying similarities subconsciously and we process, react, and respond accordingly. This is a problem for machines—as any algorithm needs the input to be in a set format, and each of these sentences vary in their structure and format. If we decided to code rules for every combination of words in any natural language; things will get very complicated, very quickly. This is where NLP enters the picture.

NLP is a subset of AI tasked with enabling machines to interact using natural languages. The domain of NLP also ensures that machines can:

- ③ Process extremely large volumes of natural language data
- ③ Derive insights and meaning from that information

Before any of this natural language processing can occur, the text needs to be normalized.

In machine learning (ML) jargon, this series of steps is called data pre-processing. The concept is to break down natural language into more manageable chunks which can then be analyzed to find relations, dependencies, and context.

What is natural language understanding (NLU)?

Considered a subtopic of NLP, the main purpose of natural language understanding is to enable machines to:

- ③ Interpret the natural language
- ③ Derive meaning
- ③ Identify context

SapienSecure utilizes NLU to translate the combined experiences of all of our subject

matter experts into a portable and on-demand electronic format for our customers. There are subtleties in the way that medical text is interpreted and standard semantic searches or keyword searches are not able to accurately capture this. We have trained our AI model to understand medical text and the context in which it's presented the same way that a physician would. Then we find applications that would benefit our customers and fine-tune the model to perform specific tasks.

For example, SapienSecure uses NLU primarily to perform Name Entity Recognition (NER) and process semantics. The NLU algorithms often operate on text that has already been standardized by text pre-processing steps we have created with our NLP models.

Going back to our weather inquiry example, it is NLU which enables the machine to understand that those three different questions have the same underlying weather forecast query. After all, different sentences can mean the same thing, and, vice versa, the same words can mean different things depending on how they are used.

Example of Natural Language Understanding:

- 🌀 The banks will be closed on Christmas Day.
- 🌀 The banks will overflow from the river during floods.

Our NLU models use what is called word sense disambiguation, which sits under the NLU umbrella assisting the machine, so it can understand how and why the two different sense's the word "bank" are used.

So, how do NLP & NLU differ?

In natural language, what is expressed (either via speech or text) is not always what is meant. NLP focuses on processing the text in a literal sense, like what was said, whereas NLU focuses on extracting the context and intent, or in other words, what was meant. Without NLU models supporting NLP, there could be damaging consequences! NLP processes text from grammar, structure, typo, and point of view—NLU guides the machine to infer the intent behind the language text.

Do we need both? In one word, yes.

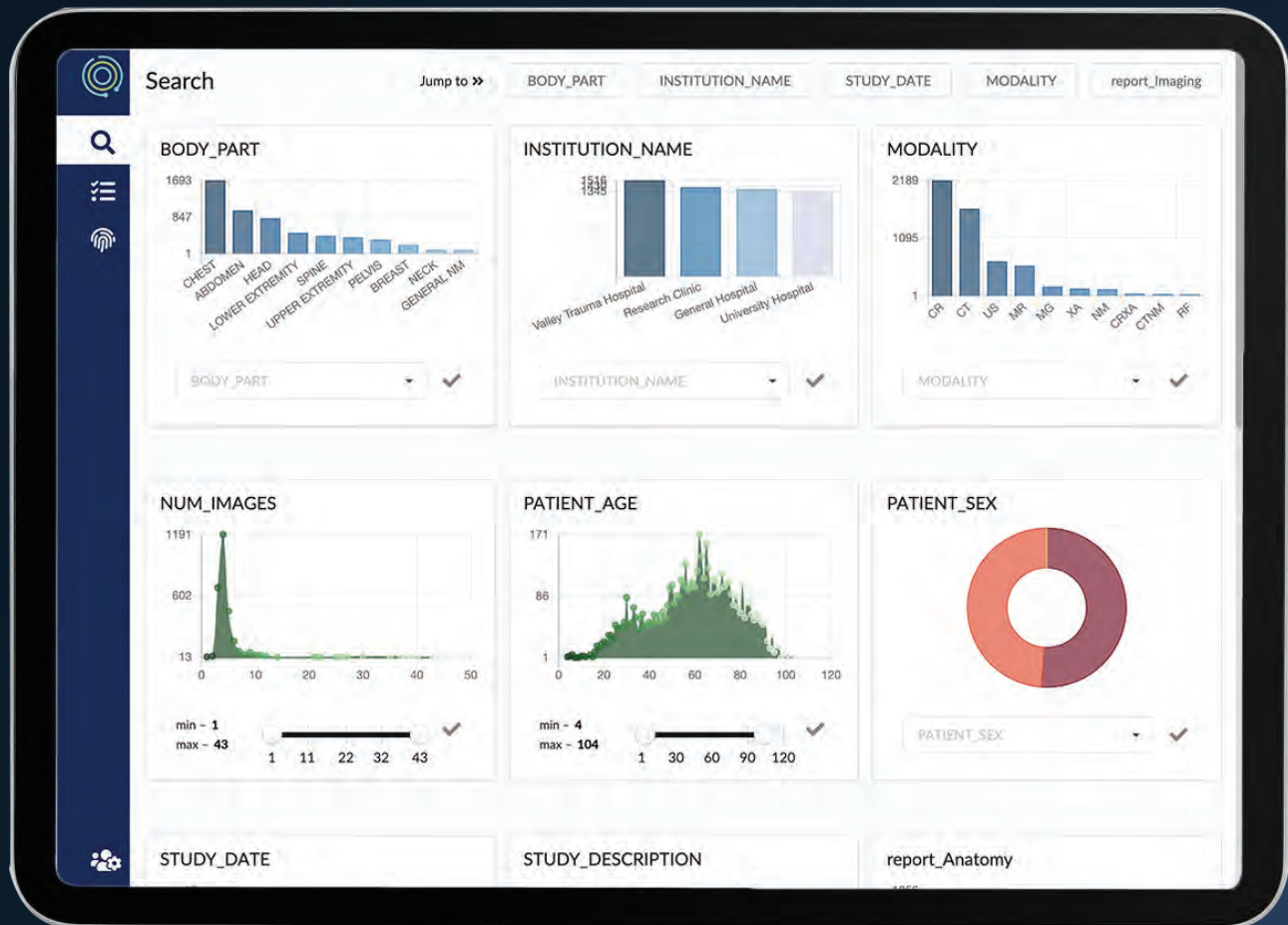
This is also what separates SapienSecure from other solutions in this space. Our NLU models were developed by practicing MDs in conjunction with our developers to ensure that our language guidance is healthcare effective. Combining NLU and NLP, the SapienSecure data extraction, classification and indexing is as robust as possible.

Some of these powerful use cases include:

- 🌀 Finding PHI in medical reports and redacting them before they are sent out to research partners.
- 🌀 Identifying the size of lung nodules as identified by radiologists in clinical reports.

Modules

DATA DISCOVERY



The Data Discovery module empowers users to explore the data within their health organization at a macro level, select the datasets of interest, de-identify them and export them to a destination.

SAPIENSECURE DASHBOARDS:

SapienDashboards is a way for health systems to display and answer specific operational and healthcare-related questions in a simple visual form factor. Dashboards refresh data on a minute-by-minute basis, and provide users with actionable information. SapienDashboards requires workshopping with customers to identify the data that matters most to their specific problem, and then a period of 1-2 weeks to create the dashboard for the customer's needs.

SAPIENSECURE BILLING:

SapienBilling is a powerful NLP-powered billing optimization module that provides medical practitioners with a view into their medical practice billings, provides workflows to ensure accurate billing code submission and utilizes NLP to recommend best billing practices.

SAPIENSECURE OCR:

SapienOCR automates the transcription of forms in health systems. SapienSecure conducts workshops with the client to determine scope, templates the forms, automates the transcription process, and operationalizes the processes in the context of the system.

SAPIENSECURE CONTAINERS:

SapienContainers is a simple AI model containerization module that makes it easy for users to implement and operationalize their AI models. Training an AI model to produce a usable prediction is one thing, but operationalizing it and containerizing it in a way that works in an organization is completely different. SapienContainers utilizes dockers, HTML endpoints, DICOM nodes and other data-flow technology to automatically route data from a health system database and through an AI model of the customer's choosing.

SAPIENSECURE AI:

SapienSecure has a skilled team of AI developers creating new AI models on a monthly basis. If an organization is interested in creating an AI model for their purposes, SapienSecure will conduct a consultancy engagement, scope out the development of the AI model, source labeling/annotation, conduct dataset curation, and train the AI model to the customer's liking. Using the SapienContainers module, the SapienSecure team can then operationalize the models and provide ongoing support.

Case Studies

AWS TURNS TO SAPIENSECURE FOR INTERNATIONAL RESEARCH COLLABORATION

AWS sought to create a pandemic-fighting AI model and turned to SapienSecure for help. Collaborators from the United States, Canada, South Korea, Italy, China, and the Middle East participated in the sharing of data, using SapienSecure, to create a Covid-19 infection model for CT scans.

Upon implementation at each of these centers, SapienSecure extracted and de-identified millions of patient records across collaborating facilities to enable AI Partners (AWS, Xtract.ai, Element AI, md.ai) to develop a state-of-the-art AI model. SapienSecure is flexible and capable of respecting multiple jurisdictional laws around privacy and security.

VANCOUVER IMAGING OPTIMIZED AND INCREASED BILLING REVENUE USING SAPIENSECURE

Vancouver Imaging approached SapienSecure with concerns that medical procedures were being billed incorrectly, leading to rejections, re-work by support staff, and potential loss of funds.

Using SapienSecure Billing, Vancouver Imaging was able to reclaim and mitigate substantial billing errors that would have never been claimed otherwise. This resulted in a decreased billing rejection rate by 30% and accounted for an increase in revenues for each physician in excess of 15%. The client's annual license fee was recovered within the first 4 months of implementation with a total annual return on investment of over 200%

SAPIENSECURE EMPOWERED PROVIDENCE HEALTHCARE DATA FOR RESEARCH AND COMMERCIALIZATION

Providence Healthcare adopted SapienSecure to build an industry leading data pipeline which transfers data to a state-of-the-art Data Lake within Amazon Web Services (AWS). This Data Lake is used for research and collaboration with industry.

Over the course of 8 months, Providence Healthcare extracted and de-identified hundreds of thousands of medical records for access within their AWS Data Lake. A dedicated internal PHC audit was conducted to test the SapienSecure platform and was conclusive in validating that zero personally identifying information existed within the outputted datasets. Pharmaceutical and Research companies were then able to access the Data Lake and conduct new research and drug discovery. Providence Healthcare generated a return of 400-500% through this work.

VANCOUVER COASTAL HEALTH TAPS SAPIENSECURE TO REDUCE WAITLISTS

Hundreds of thousands of medical imaging requests are processed by Vancouver Coastal Health annually, contributing to the large demands on human resources, equipment and operational costs, ultimately lengthening waitlist times. Vancouver Coastal Health launched a pilot project with SapienSecure in partnership with the University of British Columbia and AWS to support their wait-list reduction goals.

SapienSecure, AWS, and UBC created a natural language processing model to predict MR priority and protocol. Following the successful implementation of this pilot, VCH procured SapienSecure to operationalize and improve the SapienSecure module (MRI CANTWAIT) after projecting a 67% reduction in workforce demands.

The Future

SapienSecure empowers health systems to collect, index, discover and commercialize their medical data. The platform provides health systems with the ability to better understand and empower their own data, and analyze the data for multiple secondary purposes. SapienSecure has provided health systems with immense value, yet what if this is only the beginning?

What if publications by prominent journals, like the New England Journal of Medicine, were conducted by an automated data mining platform that required no human research assistants to collect and analyze the data? In this scenario, NLP systems would read through patient charts looking for relevant data, and including or excluding patients based on this automated search. Then the systems would convert the unstructured data (notes, reports, images, etc.), into structured tabular data that would be ingested into statistical analysis software (SciPy, R, MatLab, etc), outputting the results of a clinical hypothesis. If systems like this were adopted, constant data mining and analysis could return answers to the infinite medical questions at speeds not imaginable by current research methods.

As SapienSecure integrates more and more automation into its indexing and discovery tools, research and quality audits will become automated as well. How often did radiation doses in the radiation oncology lab surpass safety thresholds? How many patients with pneumonia did not improve with penicillin last year? Which genes are associated with short stature at my hospital? These types of simple questions will one day be answerable in milliseconds and always available based on your local health system data; instead of answerable in months, and only available if an administrator is tasked with the project as currently is the case.

Automated analytics and data processing is coming, and SapienSecure's mission is to be the first and industry leader in high quality medical data collecting, indexing, discoverability and commercialization, with the ultimate goal of changing the way discoveries are made in healthcare.



A healthcare intelligence & data discovery platform, SapienSecure® empowers your data by unlocking population data siloed in disparate databases; unifying and visualizing for the benefit of research and human care. SapienSecure enables healthcare institutions to function at a higher level of speed, accuracy, and security. Using Natural Language Processing + AI Understanding our proprietary technology supports even the most challenging imaging and unstructured data, while automating the collection, curation, and categorization of medical datasets for intelligent research and discovery; unlocking value without the need of replacing existing software or systems. Learn more at www.sapiensecure.io

Empower Your Data. Securely. Efficiently. Responsibly.

www.sapiensecure.io | 828 W 10th Ave, Suite 500, Vancouver, B.C. V5Z 1M9
