Whitepaper

Data Lake for Healthcare's Digital Transformation Journey

innovaccer

Table of Contents

Executive Summary

Shrinking reimbursements and increasing consumerism have made it critical for healthcare executives to understand and unlock the power of their organizations' data. Health systems are progressively leveraging their data to eliminate waste, enhance efficiency, and deliver upgraded patient service. For most healthcare organizations, the market demands competition on clinical, financial, and service outcomes founded on the data examined by payers, patients, and purchasers.

All these factors congregate to the chief information officer's (CIO) desk as information systems departments are tasked with implementing data in a suitable manner to meet new operational requirements. A majority of CIOs are confronted with new data warehousing overhaul efforts, as current organizational practices are hampering their adaptation to new payment models and needs. Redundant data sources propel high maintenance and operations costs for care organizations. Additionally, legacy electronic medical records (EMRs) offer cumbersome workflows and a poor user experience that decreases the efficiency of care management, patient appointments, follow-ups, and other care operations.

For successful data implementation, CIOs need to verify that their traditional data warehouses are upgraded to a powerful data lake, which serves as an efficient investment for their executive processes. To deliver value, a data lake requires specific applicability to clinical and business problems. Healthcare organizations require a complete data management system that extracts competitive outcome metrics and delivers actionable insights for them. Hospital leadership requires solutions with a strong data lake foundation that rationalize their information technology (IT) expenditures, simplify regulatory reporting for staff, and address the evolving needs of the end users. In this scenario, rationalized and aggregated data sources in a data lake save considerable IT and management costs, and facilitate a wide range of reporting requirements for health systems with customized, bolt-on solutions.

In this whitepaper, we understand the following:

- Concept of a data lake and how it offers smarter application scope than the traditional data warehouses in healthcare

- Six key accelerators in Innovaccer's Data Lake and how they help healthcare organizations in speeding up their digital transformation journey

- Estimated cost savings driven by Innovaccer's Data Lake solution, and how it helps care organizations in staying dynamic with futuristic analytical and reporting capabilities

What Is a Data Lake ?

A data lake is a consolidated repository for data that allows organizations to research, integrate, and analyze large quantities of information at once. Healthcare organizations may use it to execute clinical analytics using patient data stored in the electronic health record (EHR), or they may refine their financial forecasting by drilling into business intelligence and revenue cycle analytics using claims and billing codes.

For an organization, the most efficacious data lake is a connected care framework that center all healthcare information around individual patients and helps generate actionable information at all points of care. For health systems, it presents a cost-effective business strategy that offers simplified analytics support, accelerated implementation and speed to return on investment (ROI), substantial back-end automation, and a flexible structure that allows the organization to adapt to evolving competitive and regulatory pressures and evolve into the next generation of analytics.

Healthcare organizations CIOs, chief medical information officers (CMIOs), and chief nursing information officers (CNIOs) are shifting to these new systems to ease their struggle with analytics and adapt to new data tools. This update from traditional warehouse strategies is reinforced with the availability of new data sources and new data tools. Let's understand the benefits that Innovaccer's Data Lake brings to healthcare organizations:

- Data Lake is an integrated, holistic platform that assists all analytics and decision support requirements for health systems. It contains a comprehensive set of vertical capabilities that align perfectly with the system's core service lines and ancillary care departments.

- It aggregates patient data from a wide range of clinical and non-clinical sources into a single platform and enables health systems to unlock the true power of their data by performing analysis alongside a series of measures like expenditure and utilization for actionable insights that direct health systems to high performance in value-based metrics.

- It efficiently dissolves department silos with a horizontally-integrated framework that connects procurement, utilization, and outcomes data in an automated fashion to yield strategic insights and actionable intervention opportunities for health systems.

- It works with the next-generation analytics platform to take complete care of the health system's data governance requirements, as well as the downstream analytics needs for efficient execution of care programs.

- To meet the analytic requirements of health systems, Innovaccer's Data Lake helps to generate intuitive and engaging visuals for departmental and clinical end users, IT analysts, and database administrators. This empowers health systems with data-driven decision making.

- It helps health systems with simplified reporting by leveraging standard and custom reports to draw insights and ease the burden of reporting on their staff.

- It allows hospitals to create "unified patient records" and consume them as application programming interfaces (APIs) for network development and multiple care applications.

The Need for a Unified Data Lake

Providers and health systems have invested in a wide range of technologies over the past decade to cope with changing healthcare regulations and to improve end-user experience. Today, there is a growing focus on the cloud across businesses, and healthcare CIOs have concluded that hosting data and systems on-premise does not result in any additional benefits over cloud hosting environments.

On the other hand, adapting to the cloud-based infrastructure not only provided better uptime and performance but also allowed providers to deliver faster value to their end users and step into a more strategic role in solving problems for their patients. Hours spent in procuring, setting up and maintaining extensive infrastructure could now be saved for improving patient experience and interaction1.

As providers' health systems have moved towards a unified infrastructure and a unified system of records, the focus has increasingly shifted to data problems: how to best interoperate data across providers, how to organize data, how to improve visibility into processes with data, how to better predict trends with data, and other important needs.

Interestingly, most providers rely on separate vendors to solve for different parts of the data problem today. While having the flexibility to solve one problem at a time is important, this approach of each problem being solved by a different vendor creates an untenable norm. It eventually creates silos, increases duplicity of cost and efforts, and delivers a suboptimal ROI for providers.

Disparate healthcare data needs to be integrated into a single record which can incorporate real-time changes in the information when hosted on the cloud. A series of intelligent algorithms on the integrated record would help health systems host multiple services such as patient engagement, care management, virtual care and remote patient monitoring, provider engagement, business decision support, and process automation from a single EMR platform. It is essential for healthcare organizations to unify the broken processes and overcome data silos on an integrated data platform to weave coordinated stakeholder experiences.

Unified Data Lake Versus Traditional Data Warehouse

Data lakes and data warehouses are both used across industries for storing big data, but they are slightly different when it comes to data accessibility, usability and storage. A data lake is a vast pool of raw data, the purpose for which is not yet defined. On the other hand, a data warehouse is a repository for structured, filtered data that has already been processed for a specific purpose2.

A meaningful distinction between the two varies across industries and depends on a wide range of factors. Data lakes, for instance, were born out of the need to harness big data and benefit from the raw, granular structured and unstructured data for machine learning, while there is still a need to create data warehouses for analytics use by business users.

For the healthcare industry, data warehouses have been used for many years, but have never been particularly successful. Due to the unstructured nature of much of the data in healthcare (physicians notes, clinical data, etc.) and the need for real-time insights, data warehouses are generally not well-suited for healthcare data management.

Since data lakes allow for a combination of structured and unstructured data, they tend to be better fits for healthcare companies. They offer more utility than data warehouses, because they enhance the applicability of data to different customized, bolt-on solutions that are specifically designed to assist practitioners, patients, and care managers with their experiential challenges.

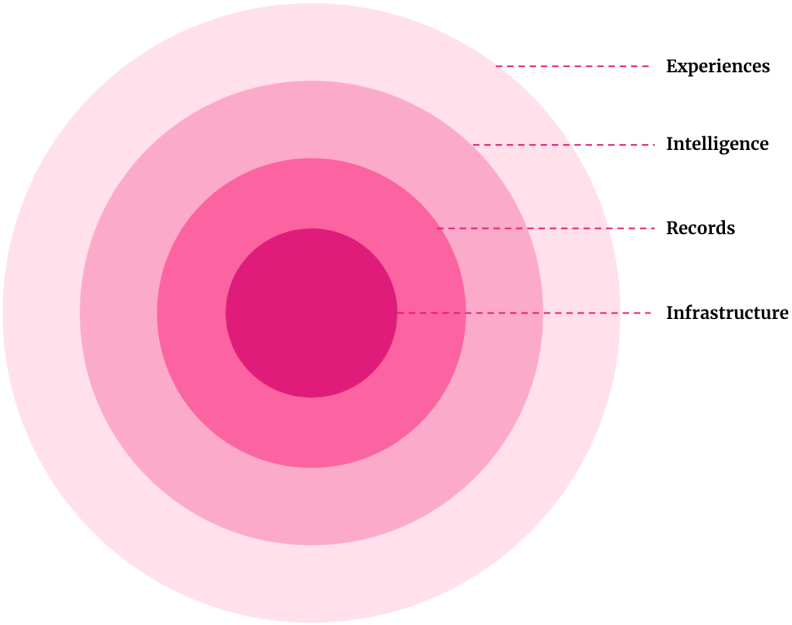
The key differences between the two are seen through the table below:

Traditional Data Warehouse	Modern Data Lake
Stores processed and refined data	Primarily store raw, unprocessed data
Designed to be structured	Allows for a combination of structured and unstructured data
Difficult and costly to manipulate	Ideal to support a variety of bolt-on use cases
More complicated to make changes	Highly accessible and quick to update

The Four Driving Layers in Healthcare's Digital Transformation Journey

With the adoption of advanced EHR generations, care organizations today have humongous amounts of clinical, social, behavioral, financial, and operational data at their disposal. They must harness the power of integrated information from these diverse, siloed pools of data. An integrated data platform redefines and monetizes each strand of care delivery with an optimized, efficient and pattern-sensitive approach, thereby paving the way to a series of impactful healthcare applications.

Healthcare's Digital Transformation Journey



In the broader view, four layers drive digital transformation in healthcare to deliver a better stakeholder experience across the care continuum:

The Infrastructure Layer: The physical servers, storage, network, and firewalls that support the applications used by various users. At the core of it, all data solutions lie in the infrastructure layer.

The Records Layer: The EMRs that providers use to record their patients' healthcare data across facilities, departments, and encounters.

The Data Transformation Layer: The data layer that helps derive operational and predictive insights from data to help users derive greater value from their data, make processes more efficient, and improve the quality of care while lowering its overall cost.

The Experience Layer: The ecosystem of end-user applications and solutions designed to streamline workflows and enhance the end-user experience for all stakeholders across the care continuum.

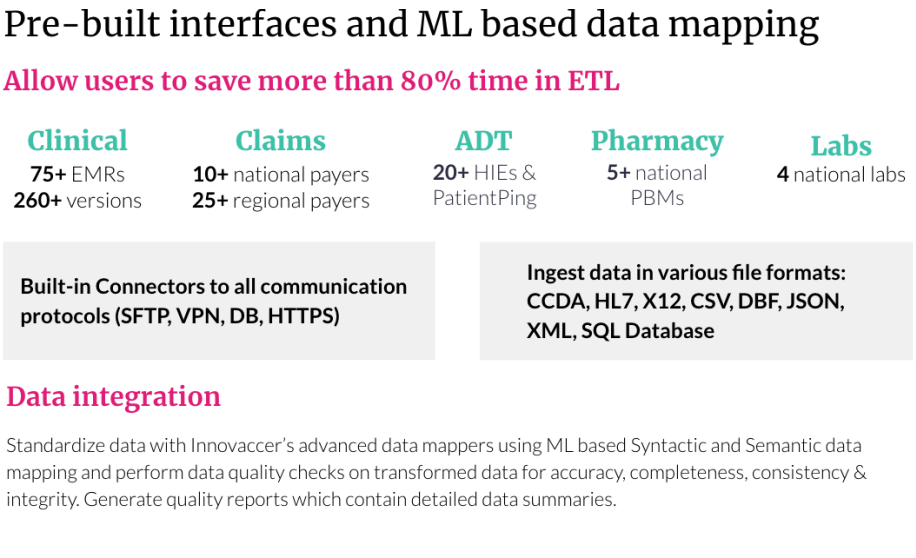
Data Transformation Layer and Innovaccer's Data Lake Key Accelerators

Innovaccer's Data Lake is foundational to an ROI framework that minimizes the health system's investments by realizing cost-saving opportunities within IT infrastructure. It is the most cost-effective data archiving solution that deploys a variety of use cases to realize productivity, efficiency, and quality impact for physicians, care coordinators, and healthcare leadership.

While most data platforms have predefined formats for data ingestion, Innovaccer works directly with the hospital's source systems to integrate data into its proprietary Data Activation Platform (DAP). Our solution provides end-to-end ownership of healthcare data management and its applications.

Our Data Lake architecture provides over 200 pre-built connectors to allow hospitals to onboard and test data in just twenty weeks. With intelligent algorithms, our platform delivers deep artificial intelligence (AI)- and machine learning (ML)- powered predictive insights to drive the hospital's operations.

1. **ETL and Interfaces:** Over the years, Innovaccer has worked with several most commonly used source data systems across health systems – EMRs, claims data, data from bedside medical devices, personal health devices, remote patient monitoring devices, human resources, finance & supply chain systems and third-party data such as social determinants of health, pollution, etc. With a deep understanding of each source system and aggregating data from multiple sources, Innovaccer has built over 85 pre-built connectors with these sources that automate most of the Extract-Transform-Load (ETL) steps and solve the data aggregation problem in record time.



2. **Data quality Validation:** Ingested data is run through DAP's DQT (Data Quality Assessment Tool) to identify gaps and errors in the ingested data and generate a "Data Quality Report." This report contains a detailed qualitative analysis of the specified dataset, including missing and duplicate values and deviations from coding standards for sixty-two data fields such as clinical, demographic, and financial codes.

EMPI and data quality tools

Continuous data quality monitoring

Data model at source systems change over time. Continuous data quality monitoring at the records level ensures that data quality issues are caught and fixed on the fly. Innovaccer offers a quality monitoring tool with the ability to track quality over time and add business rules, regex, lookups and monitor quality over all 6 of the principles of data quality.

Six principles of data quality

1. Accuracy
2. Consistency
3. Validity
4. Completeness
5. Timeliness
6. Uniqueness

80+ Coding Standards

1000+ Data Attributes

25M+ Records in Knowledge Base

6000+ Business Rules

Enterprise Master Patient Index (EMPI) and single best record

EMPI enables you to identify a unique patient from multiple data systems for quality of care assessment, payment calculations and contract performance reviews.

3. **Unified Data Model:** After the data quality check, DAP's pipelines transform the raw data, by governing the end-to-end data flow and performing the required standardizations, modifications, and other operations to ensure that only clean and accurate data is processed. Data is then mapped to DAP's master schema and stored in an integrated Data Lake.

   As Innovaccer aggregates data from multiple sources, health systems can create longitudinal records for patients from various data sources. Innovaccer's master patient index (MPI) algorithm creates a unique MPI for each patient, allows providers to identify duplicates by matching the MPIs across patients, thereby minimizing duplicate patient records.

   Processed and clean high-quality data, available in the integrated Data Lake, can be accessed through Innovaccer's library of pre-built APIs, leveraged to power custom-built applications. Later, DAP's EMPI (Enterprise Master Patient Index) engine helps uniquely identify members (patients) across disparate healthcare IT systems, imperative for any downstream applications. DAP employs our proprietary Bayesian-based flexible matching algorithm that produces more than 95% accuracy every time.

Unified Data Model

Full longitudinal record of a patient from disparate sources

Patient at the center
The data model is centered around the patient, but also supports aggregations (e.g., provider, payer) as secondary support to power up dashboards and applications.

Supports Data Lake
The UDM is optimized for every storage across the data lake (including distributed file system, ROLAP, MOLAP, and OLTPs).

Healthcare industry standards support
Broad and expanding standards support including FHIR, HL7, OMOP, HIPAA, CCD, CCPA, GDPR 1, CDISC-BRIDG, and SAMHSA.

Data level security
Data classification, archival, and retention policies, with access policies at each storage of the data lake ensuring data security across protocol access.

Extensible and backward-compatible
UDM is backward-compatible irrespective of HL7 versions, thus ensuring standard version-agnostic data model without any modifications needed to the upstream or downstream data pipelines.

Interoperability and FHIR
UDM is optimized for applications through FHIR APIs with a normalized data model. Furthermore, UDM data models at the FHIR level are compatible with interoperability tools to support two-way interoperability.

Extended genomics coverage
Support for multiple genotype and phenotype data sets.

Data Science support
Data is partitioned at multiple levels.

4. **API Management:** DAP's extensive API library includes the complete FHIR API library. Innovaccer is also working on creating a library of APIs that extends far beyond patient lists. For example, creating provider credentialing APIs helps determine providers' most current directories within a few API calls each month.

Healthcare Analytics

Full suite of analytics around five pillars: cost, quality, utilization, risk, and patient satisfaction

600+	16+	10+	250+	100+
analytical models	different type of charts	built-in dashboards	quality, utilization, and cost measures	volume and value based reports

UI-based analytical model builder to define custom measures at speed with version control

5. **Visualization:** Innovaccer's Data Activation Platform also comes with InGraph, its own business intelligence (BI) tool that accelerates the ability to derive insights from the health systems' data. The standard measures and dashboards help providers augment their business intelligence needs and deliver quick wins. Innovaccer's DAP allows health systems the flexibility to operate any SQL/AI/ML workbench of their choice to drive their analytics and power their algorithms. The data platform also allows providers to derive insights from their data and save it within the data model for further consumption by downstream applications and BI tools.

FHIR Platform features

FHIR servers & resources

HL7 FHIR®-compliant RESTful APIs (R4 and STU3 versions)

API gateways

Complete control of your platform through an intuitive administration interface including usage monitoring, traffic management, and IPs prevention

Smart on FHIR capabilities

Support for OAuth 2.0, OpenID Connect, and Smart Launch

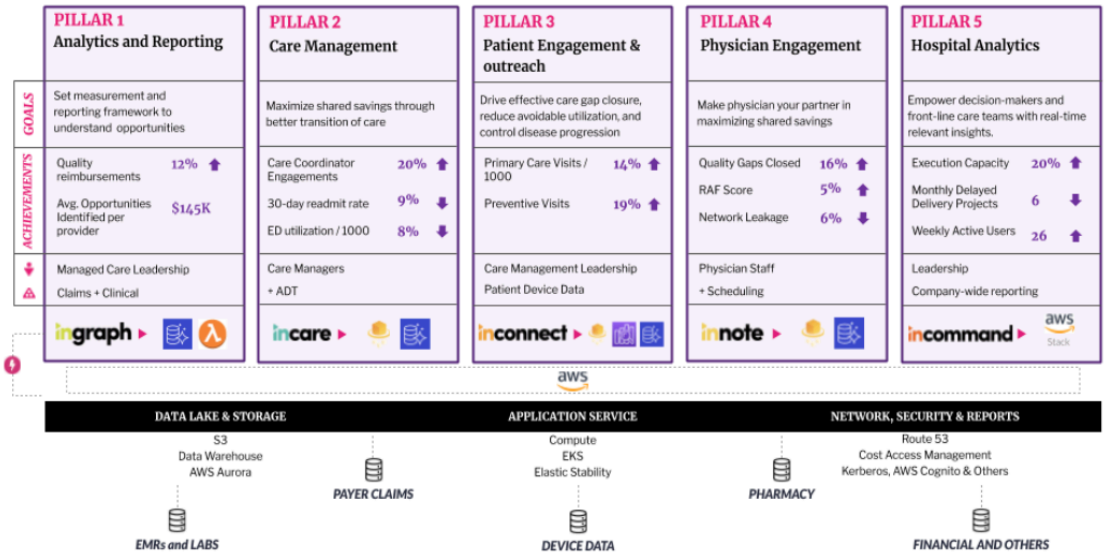
De-identification engine

PHI/PII data is de-identified for a secure, interactive developer environment

FHIR applications

FHIR applications for provider network, members, employers, and payer team members

Activating Data for Healthcare Together



Innovaccer's Data Lake: Salient Features

Data Lake supports and directs management and strategic decision making for healthcare organizations. Innovaccer partners with healthcare organizations to seamlessly integrate data from each of the clinical data sources at the hospitals, practices, and other entities that are part of their network through supported, secure, and standard data interfaces. By also including claims data, it creates a centralized and holistic single source of truth within the digital data environment for the organization.

Innovaccer then leverages our industry-leading DAP as the foundational platform on which the digital data environment is built and maintained for the healthcare organization. This digital data environment is utilized by the organization for further downstream analytics and querying through integration with existing analytical tools, utilizing Innovaccer's award-winning analytics engine, InGraph, for reporting and care plan recommendations.

Salient Features:

- Bi-directional interoperability with pre-built integration with over 65 EMRs/EHRs, more than 45 payers, at least 5 lab systems, real-time healthcare IoT Sensors, and many other health systems.

- Fully managed and customized on-demand data lake systems (Hadoop, Distributed File storages system, OLAP data warehouses, Hyperscale application databases with linear and horizontal scalability and other data storage systems)

- Analytics abilities (AI/ML framework-agnostic platform with pre-built analytics for more than 500 quality measures, 15 risk models, over 50 utilization and financial measures for payers, attribution with any triggering points, SVI, patient segmentation, loyalty, and advanced predictive analytics with an intuitive interface)

- Powerful healthcare parsers for formats such as CCDs, HL7, FHIR, and EDI

- Contextualized infrastructure and security. (Kerberised, PHI protected across data lake with centralized access system)

- Turn-key solutions using pre-built products

Innovaccer's Data Lake: Estimated Cost Savings and Value Offered

If a single, powerful solution were to replace ten different solutions trying to solve the same problem together, it would not just bring greater dynamicity and flexibility to adapt new technologies, but also reduce the cost of the investment.

Consider a healthcare organization that leverages two or three different warehouses to store its clinical, financial, administrative and other relevant data from various sources. The organization ends up paying significant sums for data duplicacy, duplicate interfaces and different licenses. The benefits don't end there; the number of people hired to manage these different warehouses can also be optimized if a unified, powerful Data Lake replaces these warehouses.

The table below shows the benefits that Innovaccer's enterprise platform on Azure to build a Data Lake offers, versus a native build on Azure from scratch. A series of accelerators like ETL, Data Model, Quality Monitoring, Outbound Feeds, Reporting, and Ad-Hoc Querying help the organization in saving their precious dollars. The table shows an example with a sample number of units and respective estimated savings; however, these numbers vary across organizations.

Accelerator	Description	Unit	# of Units	$ saved per unit	Total $ saving
ETL	Pre-built templates and UI enabled healthcare ETL saves time	No of integrations	70 (not in pop health)	$3,000	$210,000
Data Model	Data model build out reduction with Innovaccer data model	No of data model categories (clinical, claims, inventory….)	10	$6,000	$60,000
Quality Monitoring	Innovaccer quality monitoring infrastructure makes it faster to setup	No of integrations	70	$2,000	$105,000
Outbound feeds	Build one scheduled job versus a few to enable outbound feeds	No of outbound feeds	80	$4,000	$320,000
Reporting	Library of widgets and healthcare formulas reduces reporting build out	# of reports	400-500	$1,500	$675,000
Ad-Hoc Querying	Time saved to query one model versus tens	# of ad-hoc queries per year	30,000	$40	$1,200,000 / yr
				Total	**$1.4M + $1.2M / yr**

Figure: Estimated Savings with Accelerators versus Native Build

Value Offered:

Innovaccer's Data Lake is assisting healthcare organizations undergo a digital transformation to transition from on-premise to the cloud to save on costs and improve performance.

600+ pre-built analytics
Activated data leveraging highly customizable and extensible analytics and interactive dashboards, including pre-built views and measures engine

65+ Pre-built, rich integrations
One-click integrations Patient Ping, Aunt Bertha, and 65+ pre-built EMR connectors, enabling rapid speed to value

UDM / UPR
Patient centered data model with optimizations for various business scenarios

De-identification engine
PHI/PII data is de-identified for a secure, interactive developer environment

AI/ML/BI capabilities
Activated data leveraging highly customizable and extensible analytics and interactive dashboards, including pre-built views and measures engine

FHIR APIs with Smart on FHIR
Extensive support for R4 FHIR resources to address all of your interoperability challenges, as well as support for OAuth 2.0, OpenID Connect, and Smart Launch

BI-directional Interoperability
Activated data leveraging highly customizable and extensible analytics and interactive dashboards, including pre-built views and measures engine

Data Quality & Security
Extensive support for R4 FHIR resources to address all of your interoperability challenges, as well as support for OAuth 2.0, OpenID Connect, and Smart Launch

Figure: Key Features of Data Lake

It enables a powerful platform that will empower healthcare organizations to effectively manage their data and provide a complete view of individual patients' health, as well as the entire patient population in aggregation. Innovaccer has emerged as the leading healthcare data platform over the last five years, offering end-to-end solutions and services starting from aggregating data from various different sources, activating that data for healthcare use cases, all the way to enabling patient, clinical, and administrative workflows.

About Innovaccer

Innovaccer, Inc. is a leading San Francisco-based healthcare technology company committed to helping healthcare care as one. The company is recognized as a Best in KLAS vendor for 2021 in Population Health Management and #1 customer-rated vendor by Blackbook. Using its Data Activation Platform, Innovaccer unifies patient records and leverages artificial intelligence and analytics to automate routine workflows and facilitate whole-person care. Its solutions have been deployed across more than 1,000 locations in the U.S., enabling more than 37,000 providers to transform care delivery and work collaboratively with payers, employers and life sciences companies. By using the connected care framework, Innovaccer has helped healthcare organizations unify records for more than 24 million people and generate more than $600M in savings for the healthcare ecosystem.

535 Mission Street Floor 14th, San Francisco, CA 94105.
innovaccer.com

+1 415 504 3851

team@innovaccer.com