Informatica
from Salesforce

White Paper

Trusted Data for Healthcare AI and Analytics

Where data & AI come to LIFE

Contents

The Foundation for Responsible, Scalable Intelligence

Introduction

Artificial intelligence has the potential to predict disease earlier, personalize treatment paths, and optimize operations at scale — but most healthcare organizations still struggle to extract meaningful value because their data foundation isn't ready. The challenge isn't in the sophistication of models or algorithms — it's in the quality and trust of the data powering them.

Untrusted, fragmented, complex, and siloed data prevent AI from scaling beyond proof-of-concept pilots. Healthcare organizations must build a foundation of trusted data to generate accurate, governed, contextual, and explainable insights that drive innovation and success. In this white paper, we outline how organizations can prepare their data for AI and analytics through modern, automated, and compliance-by-design approaches, providing a clear roadmap for moving from data chaos to intelligent experiences.

It presents a practical, eight-part roadmap for healthcare organizations to move from fragmented, unreliable data to trusted intelligence that powers scalable, responsible AI.

Part One examines why healthcare AI often stalls at scale. Part Two defines the five pillars of trusted data needed to overcome these barriers. Part Three introduces a step-by-step framework for operationalizing those pillars, while Part Four quantifies the ROI and measurable impact of data readiness.

Part Five explores how to choose the right data partner, and Part Six provides an implementation roadmap. Part Seven dives into Informatica's Intelligent Management Data Cloud (IDMC), while Part Eight details practical AI use cases in healthcare.

Part One: The Challenge

Healthcare AI Is Stalling at Scale

In a recent industry analysis, 80% of healthcare AI initiatives failed to scale due to poor data readiness.

Five Critical Barriers to AI Success

1. Data Silos Create Fragmentation

Healthcare data resides in disconnected systems — electronic health records, claims databases, laboratory information systems, genomic data, wearable devices, imaging archives, and external real-world evidence repositories — that rarely communicate. This fragmentation creates a major bottleneck when AI models must draw insights across multiple data sources. For instance, a predictive readmission model may need to integrate EHR clinical data, claims history, social determinants of health, and behavioral data — all stored in separate systems with inconsistent formats and governance standards.

2. Poor Data Quality Undermines Model Performance

Even well-designed AI models underperform when trained on low-quality data. Inconsistent data formats, duplicate patient records, missing values, unmapped clinical terminologies, and outdated information create cascading issues that degrade model accuracy and reliability. Studies show that organizations with poor data quality experience 2-3x worse model performance and slower production deployment timelines.

3. Manual Data Preparation Consumes 70-80% of Project Time

Data science and analytics teams spend most of their time not on model development or innovation, but on data discovery, cleaning, validation, and preparation. This manual, repetitive work is not only inefficient — it's error-prone and prevents teams from focusing on higher-value analytics and AI development. When 70-80% of a project's effort goes to data preparation, innovation stalls.

4. Lack of Data Governance Increases Risk

Without transparent data lineage, security controls, and explainability, healthcare organizations can't confidently deploy AI in regulated environments. Compliance audits require documentation of data origin, transformation, and usage in decision-making. When governance is manual or incomplete, audits become costly, compliance is difficult to prove, and trust in AI outputs erodes. Additionally, regulations like HIPAA, GDPR, and emerging AI governance frameworks (such as FDA guidance on AI/ML in medical devices) require organizations to demonstrate robust data governance and model explainability.

5. Siloed Analytics and BI Teams Cannot Collaborate Effectively

Different business units — clinical operations, revenue cycle, quality and safety, and research teams — often maintain separate data environments, definitions, and analytics tools. This prevents the organization from maintaining a single source of truth and leads to conflicting insights. For example, a readmission rate calculated by clinical analysts may differ from the rate in the finance system, creating confusion and undermining confidence in AI-driven insights.

The Business Impact

The consequences of these barriers are significant:

- **Delayed Time-to-Value:** AI projects take 2-3x longer to progress from pilot to production.
- **Lower Model Performance:** Poor data quality directly reduces prediction accuracy by 30-40%.
- **Increased Compliance Risk:** Manual governance processes make audits more expensive and increase exposure to regulatory risk.
- **Wasted Resources:** Teams spend more time on data preparation than strategic analytics.
- **Lost Competitive Advantage:** Organizations unable to scale AI fall behind competitors who are succeeding.

Key Takeaway

Despite widespread experimentation, most healthcare AI projects fail to scale. The root causes include data silos, poor quality, manual preparation, weak governance, and fragmented analytics efforts. These barriers lead to longer deployment times, lower model accuracy, compliance risks, and wasted resources. To move beyond proof-of-concept AI, organizations must address these foundational data challenges and establish systems that ensure accuracy, transparency, and collaboration.

Part Two: Understanding Trusted Data for AI

What is "Trusted" Data?

Trusted data is reliable, comprehensive, and transparent — the foundation of responsible AI. It ensures every prediction, recommendation, and clinical decision can be traced, verified, and explained. With trusted data, healthcare organizations can move from reactive reporting to predictive, data-driven intelligence.

Informatica defines trusted data for healthcare AI through five interdependent pillars:

1. Accuracy

Data must be cleansed, standardized, and validated before it reaches AI models. This means detecting and correcting inconsistencies, intelligently handling missing values, resolving duplicate records, mapping clinical terminologies to standard vocabularies (such as SNOMED CT, ICD-10, and LOINC), and ensuring that data adheres to expected formats and business rules. Accurate data forms the foundation for any AI initiative — without it, even the most advanced models will produce unreliable outputs.

2. Integration

AI models need a 360-degree view of patients, providers, and products. Achieving this requires seamless integration across diverse data sources. Integration provides essential context, allowing AI to identify relationships and patterns that isolated datasets cannot reveal.

3. Security

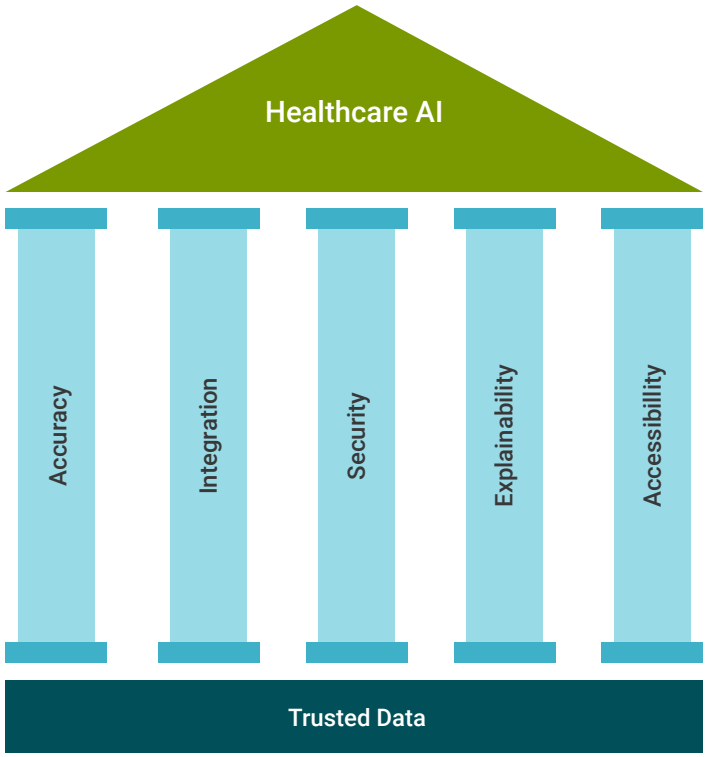
Healthcare data is sensitive and highly regulated. Trusted data platforms must enforce robust security controls — including encryption, access management, audit trails — while meeting requirements under HIPAA, GDPR, state privacy laws, and life sciences regulations like GxP and 21 CFR Part 11. As AI governance frameworks evolve, platforms should also support explainability, bias detection, and automated policy enforcement to align with emerging regulatory and ethical standards.

4. Explainability

AI models in healthcare must be interpretable, auditable, and transparent. Maintaining complete data lineage — including source, transformations, applied business rules, and model flow — enables organizations to understand and explain every decision. Explainability helps data stewards validate model behavior, supports regulators in audits, and builds clinician trust. For example, a physician reviewing an AI-generated readmission risk score should understand which data elements influenced that prediction.

5. Accessibility

Trusted data must be accessible in real-time or near-real-time for analytics and AI pipelines. Modern data architectures — such as Data Fabric, Data Mesh, and Lakehouse — enable scalable, efficient data flow without bottlenecks. Healthcare organizations must ensure data reaches the right teams and tools when needed, whether for population health analytics, AI model retraining, or clinical decision support.



Key Takeaways

Together, these five pillars — accuracy, integration, security, explainability, and accessibility — form the foundation of trusted data. Each strengthens the others: accuracy ensures confidence, integration provides context, and security and explainability enable compliance and clinician trust.

Part Three: Data Readiness Framework

From Raw Data to Trusted Intelligence

For AI to operate at scale, healthcare organizations need a clear, repeatable process for transforming raw data from source systems into trusted intelligence that powers AI and analytics.

Informatica's Healthcare AI Data Readiness Framework provides this roadmap.

Step 1: Connect Every Source Seamlessly

Objective: Identify and connect all relevant data sources, then ingest data reliably into a centralized platform.

The first step is comprehensive data discovery. Healthcare organizations must map all data sources that can inform AI models and analytics.

Once sources are identified, modern data integration tools streamline the ingestion process, handling different formats (FHIR, HL7, proprietary APIs), transformation rules, and incremental updates. Intelligent, low-code or no-code ingestion tools reduce manual effort, enabling organizations to move any data, of any type, at any latency — without overwhelming IT teams.

Step 2: Automate Data Quality at Scale

Objective: Automate data quality and validation to eliminate errors and inconsistencies.

Once data is ingested into a centralized platform, quality rules can be applied to eliminate errors and inconsistencies. Raw healthcare data is inherently messy — patient names may be spelled inconsistently, diagnoses may use multiple coding systems, lab values may have data-entry errors, and timestamps (such as encounter or procedure time records) may be formatted differently across systems. Manual cleaning is slow and error-prone.

Modern AI-powered data quality tools use machine learning to detect anomalies, recommend corrections, and standardize data automatically. These tools can identify duplicate records, validate entries against business rules, map clinical terminologies to standard vocabularies, and flag outliers — reducing manual data prep time by 60-80% and improving organization-wide consistency.

Step 3: Build a Single Source of Truth

Objective: Establish unified 360-degree views of patients, providers, and products with consistent governance.

Data mastering is key to creating a single source of truth. Patient Master Data Management (MDM), for example, resolves duplicates, generates a unified patient identifier, and maintains a trusted longitudinal patient record that analytics and AI models can rely on. Provider and product mastering apply the same principle — creating golden records for enterprise-wide use.

Governance ensures consistent and secure data use. It defines ownership, quality standards, access controls, and lineage documentation. When automated and centralized, governance enables every department to operate with the same definitions — accelerating compliance, collaboration, and trust.

Step 4: Embed Trust and Compliance by Design

Objective: Apply automated controls to meet HIPAA, GxP, 21 CFR Part 11, and emerging AI governance standards.

Security and compliance must be designed from the start — not added later. Data platforms must enforce encryption, role-based access controls, masking, and comprehensive audit trails by default.

For healthcare and life sciences organizations, compliance with HIPAA, GxP, and 21 CFR Part 11 is non-negotiable. These frameworks demand validated systems, reproducible processes, and auditable data flows. Modern cloud platforms automate many of these requirements through configuration — reducing manual effort and minimizing compliance risk.

Step 5: Deliver Real-Time, Explainable Intelligence

Objective: Deliver trusted, contextual data to AI/ML models and analytics-ready pipelines for real-time decision support.

The final step is to operationalize AI responsibly. Healthcare organizations must first build and validate AI/ML models on high-quality, trusted data to ensure explainability, accuracy, and compliance. MLOps practices then automate deployment, monitoring, and retraining, enabling safe, scalable AI adoption.

Clean, governed, integrated data then flows into analytics and AI platforms in real-time or near real-time. This enables powerful use cases like predictive readmission models that score patients daily, dynamic risk adjustment for payer pricing, real-time fraud detection engines analyzing provider behavior, and clinical decision support systems that surface evidence-based insights at the point of care.

The framework also establishes a continuous feedback loop: as AI models generate outcomes, those outcomes feed back into the platform for retraining and ongoing improvement, creating a self-reinforcing engine of intelligence.

Key Takeaway

Informatica's five-step Data Readiness Framework provides a blueprint for turning raw data into trusted intelligence. It begins with connecting every data source, automating quality at scale, and building a single source of truth through mastering and governance. Trust and compliance must be embedded by design, followed by delivering explainable, real-time intelligence to AI systems. This framework creates a continuous feedback loop that refines data and models over time — enabling AI to scale safely, efficiently, and confidently.

Part Four: ROI and Business Impact

Quantifying the Value of Trusted Data

Following the 5-step Data Readiness Framework delivers measurable ROI by accelerating timelines, reducing costs, improving data quality, and minimizing compliance risk.

Impact Area	Data Readiness		Business
	BEFORE	AFTER	
Data Prep Effort	70%	20%	3-5x faster AI model deployment; teams focus on innovation instead of manual prep
Data Quality (Duplicate/Incomplete Records)	25%	<5%	30-40% improvement in model accuracy; better clinical decisions; improved patient outcomes
Time to Production	8 Weeks	4 Weeks	$500K-$2M savings per project from faster deployment and time-to-value
Compliance Audit Prep (Time Spent On Manual Documentation)	200 Hours	50 Hours	75% reduction in audit burden; lower compliance risk; faster audit resolution
Cost Per Project	High Manual Effort + Rework	Lower Effort with Automation	$3.85M annual savings at enterprise scale (across multiple projects annually)
Model Performance	Baseline (Poor Data Quality)	+30-40% Accuracy Improvement	Better predictions; higher clinician adoption; improved patient outcomes

ROI Calculation Methodology

To quantify the ROI of trusted healthcare data management, we use the following formula:

Total ROI = (Labor Hours Reduced × Hourly Labor Cost) + (Compliance Risk Avoided × Compliance Cost Factor) + (Faster Deployment × Time-to-Value Acceleration)

Example:

A healthcare system with 10 concurrent analytics projects, each requiring 2,400 hours of data prep at $75/hour, reduces prep time by 50% through trusted data investment.

Labor Savings Calculation:

Annual labor savings = 10 projects × 2,400 hours per project × 0.5 time reduction × $75/ hour labor rate = **$900,000**.

Adding Incremental Benefits From:

- Compliance risk reduction: $600K-$900K annually
- Accelerated deployment and faster go-live: $600K-$1M annually

Estimated Total Annual Savings: $2.1-2.8M.

Key Takeaway

Trusted data delivers a measurable ROI by reducing manual effort, accelerating deployment, and improving accuracy. Organizations that follow the 5-step Data Readiness Framework can redirect resources toward innovation and realize tangible financial and clinical benefits.

Part Five: Choosing the Right Partner

Key Evaluation Criteria

As you evaluate data management platforms for healthcare AI, consider these critical factors:

Healthcare-Specific Expertise: Select a partner with deep understanding of healthcare data complexity, regulatory requirements, and clinical workflows. Prioritize vendors with proven healthcare implementations and established compliance certifications.

Cloud-Native Architecture: Choose vendors that provide modern, cloud-based platforms that scale elastically, support hybrid and multi-cloud environments, and integrate with leading analytics platforms such as Snowflake, Databricks, AWS, Azure, and GCP. Avoid legacy on-premises solutions that struggle with healthcare-scale data, lack real-time processing, or fail to support unstructured data. Ensure the platform supports modern architectures to enable advanced analytics and AI consumption.

Cloud Modernization: Beyond cloud adoption, prioritize full cloud modernization — rethinking how data is stored, processed, and accessed to unlock agility and innovation. Modern data platforms can process structured, semi-structured, and unstructured data in real time from EHRs, IoT devices, imaging systems, and clinical notes. They also support event-driven architectures and real-time analytics for use cases like predictive alerts, AI-assisted decision support, and proactive patient engagement. By modernizing both infrastructure and integration pipelines, organizations can future-proof their data foundation for AI-driven transformation.

AI-Powered Automation: Manual data management slows innovation and introduces errors. Select vendors that use AI/ML to automate data quality, anomaly detection, master data matching, and governance — dramatically reducing manual effort and improving accuracy.

Transparent Lineage and Explainability: Transparency is essential for regulatory compliance and clinician trust. Ensure the platform provides end-to-end data lineage and supports model explainability for audits, regulatory submissions, and confidence in AI-driven insights.

Compliance and Security by Design: Security and compliance cannot be an afterthought. Evaluate each vendor's architecture, certifications (HIPAA, GDPR, SOC 2), and compliance capabilities such as audit trails, encryption, access controls, and data masking).

Partner Ecosystem: Select vendors with strong partnerships and proven interoperability with your existing technology stack. A robust ecosystem accelerates implementation, enhances flexibility, and reduces integration risk.

Key Takeaway

Selecting the right data partner is essential for scaling trusted AI. The ideal partner offers deep healthcare expertise, a cloud-native and modernized architecture, AI-powered automation, transparent lineage, and compliance-by-design. Strong ecosystem partnerships and interoperability with major cloud and analytics platforms ensure seamless integration. Informatica's IDMC meets these criteria, empowering healthcare organizations to operationalize trusted data efficiently, securely, and at scale.

Part Six: Implementation Roadmap

Moving from Insight to Action

Successful trusted data initiatives follow a structured implementation roadmap. While each organization's path is unique, the stages are consistent:

Phase 1: Assess and Align (Weeks 1-4)

- **Current state assessment:** Inventory data sources, identify quality issues, document governance gaps.
- **Use case prioritization:** Select 2-3 high-impact AI or analytics projects for initial pilots.
- **Technical design:** Define architecture, select cloud platform, and determine integration approach.
- **Business case refinement:** Validate ROI assumptions and align success metrics with organizational priorities.

Phase 2: Build the Foundation (Weeks 5-16)

- **Cloud setup:** Deploy the cloud data management platform within your environment.
- **Source system integration:** Connect to key systems and data sources.
- **Data quality rules:** Define and implement validation rules and quality benchmarks.
- **Master data setup:** Develop patient, provider, and product master data processes to establish single sources of truth.
- **Security and compliance:** Implement HIPAA, GDPR, and other relevant regulatory controls.

Phase 3: Pilot and Optimize (Weeks 17-28)

- **Initial deployment:** Launch the first AI/analytics use case using trusted, validated data.
- **Monitoring and optimization:** Track key metrics — data quality, latency, and model performance — to measure impact.
- **Process refinement:** Document lessons learned, refine business rules, and adjust governance policies.
- **Stakeholder expansion:** Extend participation in additional business units and new AI/analytics use cases.

Phase 4: Scale Enterprise-Wide (Weeks 29-52)

- **Multi-project deployment:** Execute additional AI and analytics projects using the same trusted data foundation.
- **Continuous improvement:** Enhance data quality rules, expand governance frameworks, optimize system performance.
- **Organizational enablement:** Train teams, establish data and AI Centers of Excellence (CoEs), and share best practices.
- **Strategic expansion:** Identify next-generation AI use cases made possible by the trusted data foundation.

Phase 5: Continuously Innovate (Ongoing)

- **Model retraining:** Implement automated retraining pipelines for continuous model improvement.
- **New data source integration:** Incorporate new and emerging data types (e.g., IoT, genomics, patient-generated data).
- **Regulatory readiness:** Stay aligned with evolving data privacy and AI governance regulations.
- **Competitive advantage:** Use trusted data as a long-term differentiator for innovation and growth.

Key Takeaway

A structured, phased roadmap is essential for successful trusted data deployment. It starts with assessment and alignment, builds the foundational architecture, pilots and optimizes early use cases, and then scales enterprise-wide. Finally, organizations continuously innovate — transforming data governance and AI readiness from a one-time initiative into an ongoing strategic capability that evolves with new data types, regulations, and AI opportunities.

Part Seven: Informatica's Solution for Healthcare Trusted Data

The Intelligent Data Management Cloud (IDMC)

Informatica's IDMC is a comprehensive cloud-native platform designed specifically for healthcare organizations to create trusted data for AI and analytics. IDMC automates the entire data readiness workflow — ingestion, mastering, cleansing, governance, and activation — powered by CLAIRE, Informatica's AI engine.

Key Capabilities

Automated Data Integration and Ingestion

IDMC connects seamlessly to all major healthcare systems and data sources — including EHRs, claims systems, lab systems, wearables, genomic databases, and external APIs. Automated ingestion handles format translation, incremental updates, and scheduling, reducing manual ETL effort by building AI-ready data pipelines in minutes via low-code/no-code capabilities.

AI-Powered Data Quality and Cleansing

CLAIRE, Informatica's AI-powered metadata engine, detects anomalies, recommends corrections, and standardizes data at scale. Machine learning algorithms identify duplicate records, validate against business rules, map clinical terminologies to standard vocabularies, and flag outliers — reducing manual data prep time by 60-80%.

Patient, Provider, and Product Mastering

IDMC creates unified golden records that eliminate data conflicts and establish trusted, 360-degree views of patients, providers, and products. Centralized master data governance ensures consistency across systems and builds confidence in analytics and AI outputs.

Comprehensive Data Governance and Lineage

In addition to mastering, IDMC provides full data lineage — documenting data origins, transformations, policies, and usage. This transparency supports explainability, accelerates compliance audits, and reduces audit preparation time by up to 75%.

Security and Compliance Automation

Built-in security and compliance controls automatically enforce HIPAA safeguards, GDPR compliance, GxP validation, and 21 CFR Part 11 requirements. Role-based access control, encryption, audit trails, and compliance reporting are automated, reducing manual compliance burden.

Real-Time Data Activation

IDMC delivers clean, governed data in real-time or near-real-time to analytics platforms, AI pipelines, and clinical systems. Its modern architecture supports batch, streaming, and API-based data delivery to meet a wide range of use case and latency requirements.

IDMC Impact in Numbers

Organizations using IDMC report measurable outcomes, including:

- **2-4x faster AI model deployment** from pilot to production
- **30-40% improvement in model accuracy** through better data quality
- **75% reduction in compliance audit tim**e through automated lineage and governance
- **60-80% reduction in manual data prep effort**, freeing teams to focus on innovation
- **3-5x ROI** on data governance investments within 18-24 months

Real-World Case Studies

Global Biopharma: Clinical Trial Acceleration

A global biopharma company faced lengthy clinical trial cycles due to manual data harmonization across multiple trial sites, patient populations, and data systems. After implementing IDMC to automate patient, site, and lab data harmonization, the company achieved:

- 25% reduction in trial cycle time
- Faster database lock enabling faster analysis and publication
- $4.8M annual ROI from accelerated trial timelines and faster regulatory submissions

Healthcare Provider: Readmission Prediction

A large health system used IDMC to deploy an AI model predicting patient readmissions with unified clinical and claims data. With trusted data in place, the model achieved:

- 35% improvement in prediction accuracy compared to initial pilots
- 2.4x faster model deployment (from 12 weeks to 5 weeks)
- $1.2M annual savings from prevented readmissions (via targeted interventions)
- Reduction of compliance audit time from 160 hours to 45 hours annually

Health Plan: Fraud Detection

A national payer used IDMC to implement an AI-powered fraud detection system using trusted claims and referral data, resulting in:

- $18M in fraudulent claims detected in year one
- 40% reduction in manual review time for investigators
- 50% improvement in detection accuracy through comprehensive data integration

Key Takeaway

Informatica's IDMC delivers the automation, scalability, and governance healthcare AI requires. Powered by CLAIRE AI, IDMC automates ingestion, cleansing, mastering, and compliance. It enables organizations to reduce manual effort by up to 80%, accelerate AI deployment by 3-4x, and achieve significant ROI. Real-world success stories from biopharma, providers, and payers demonstrate IDMC's ability to reduce trial timelines, improve predictions, and enhance fraud detection.

Part Eight: AI Use Cases Across Healthcare

Trusted data enables healthcare organizations to move from reactive reporting to predictive intelligence. Here are proven use cases across providers, payers, and life sciences:

Providers: Clinical Excellence and Operations

Predictive Readmission Models use patient history, clinical data, social determinants, and behavioral data to identify high-risk patients before discharge. Interventions based on these predictions — early follow-up, care coordination, medication reconciliation — can reduce preventable readmissions by 15-20%, directly improving patient outcomes and reducing costs.

Precision Medicine and Treatment Personalization integrate EHR data with genomic information, wearable data, and treatment outcomes to recommend personalized treatment plans. This enables clinicians to make evidence-based decisions tailored to individual patients, improving efficacy, and reducing adverse events.

Operational Optimization employs integrated EHR, scheduling, supply chain, and financial data to optimize staffing, equipment utilization, and resource allocation. Predictive models can forecast demand, recommend staffing levels, and identify inefficiencies — reducing operational costs by 10-15%.

Quality and Safety models detect patterns that correlate with adverse events, infections, or patient safety risks, enabling proactive interventions. Trusted data lineage ensures that all quality insights can be audited and explained to clinicians and regulators.

Payers: Risk Intelligence and Cost Management

Risk Scoring and Pricing takes comprehensive member data — claims history, clinical conditions, social determinants, behavioral patterns — and uses it to accurately assess member risk. Better risk adjustment improves pricing accuracy, enables targeted interventions, and reduces adverse selection risk.

Fraud, Waste, and Abuse Detection implements integrated provider claims, referral, and outcome data to identify suspicious patterns. AI models trained on trusted data can detect billing anomalies, inappropriate referrals, and unnecessary procedures, recovering millions in improper payments annually.

Prior Authorization Optimization takes advantage of historical approval data, clinical evidence, and outcome data to automate routine authorization decisions. This reduces administrative burden, improves provider experience, and ensures appropriate care reaches patients faster.

Care Management Targeting uses integrated data to identify members who would benefit most from care management interventions — such as complex case management, disease management, or behavioral health support. Predictive targeting improves intervention effectiveness and ROI.

Life Sciences: Evidence Generation and Regulatory Success

Clinical Trial Acceleration utilizes harmonized patient data, genomic information, and outcome data to identify eligible patients faster, reduce trial timelines, and accelerate regulatory submissions. Trusted data reduces the time from database lock to analysis — a critical path item in trial completion.

Real-World Evidence Generation integrates claims, EHR, and external data sources to demonstrate product effectiveness in real-world settings. This supports regulatory submissions, payer negotiations, and medical evidence generation.

Drug Safety and Pharmacovigilance uses integrated adverse event data, claims data, and medical literature to detect safety signals faster. Trusted data ensures that safety insights can be traced, verified, and communicated to regulators with full transparency.

Biomarker and Patient Stratification Research applies genomic, clinical, and outcome data to identify patient populations most likely to benefit from new therapies. This informs trial design, enhances regulatory submissions, and accelerates precision medicine development.

Key Takeaway

Trusted data powers transformative AI use cases across the healthcare ecosystem

- Providers use it for readmission prediction, personalized medicine, and operational optimization.
- Payers leverage it for risk scoring, fraud detection, and prior authorization automation.
- Life sciences organizations rely on it for faster trials, real-world evidence generation, and pharmacovigilance.

Across these domains, trusted data enables a shift from reactive analytics to predictive, explainable intelligence that improves outcomes and efficiency.

Conclusion: Building the Future of Healthcare AI

AI holds immense potential to improve patient outcomes, personalize care, and optimize operations. But realizing that potential depends on one essential element: a foundation of trusted data.

Healthcare leaders who prioritize data readiness with Informatica's IDMC will gain a decisive advantage. They'll deploy AI faster, achieve higher accuracy, reduce compliance risk, and deliver better outcomes for patients and stronger financial performance.

The path to trusted data is clear — and achievable today. Modern platforms like IDMC accelerate this transformation by automating complexity, reducing manual effort, and empowering healthcare organizations to move confidently from data chaos to AI at scale.

The question isn't whether to invest in trusted data. It's how quickly you can begin.

By investing in trusted data now, healthcare organizations can turn fragmented information into governed intelligence — driving safer, smarter, and more equitable care. With Informatica's IDMC, the future of healthcare AI isn't just possible, it's within reach. Contact your Informatica account representative to learn more.

Executive Summary

The Foundation for Responsible, Scalable Intelligence

Artificial intelligence promises to revolutionize patient care, accelerate research, and optimize operations — but 80% of healthcare AI initiatives fail to scale. The challenge isn't model sophistication, it's data readiness. Fragmented, inconsistent, and siloed data undermine model accuracy, delay deployments, and increase compliance risk.

The Path Forward: Building the Trusted Data Foundation

To unlock AI's potential, healthcare organizations must establish a foundation of trusted data — data that is accurate, integrated, secure, explainable, and accessible. These five pillars ensure data is reliable, governed, and available in real-time, forming the backbone of responsible, scalable intelligence. Together, they enable 3-5x faster AI deployment, improved compliance, and stronger clinician trust.

A Proven Framework for Data Readiness

Informatica's Healthcare AI Data Readiness Framework provides a repeatable roadmap to transform raw, fragmented data into trusted intelligence:

1. **Connect Every Source Seamlessly:** unify data across clinical, operational, and financial systems.
2. **Automate Data Quality at Scale:** use AI to detect and correct inconsistencies.
3. **Build a Single Source of Truth:** establish governed master data for patients, providers, and products.
4. **Embed Trust and Compliance by Design:** automate security and audit controls.
5. **Deliver Real-Time, Explainable Intelligence:** power AI and analytics with reliable, contextual data.

Quantifiable ROI and Business Impact

Organizations that modernize their data infrastructure achieve:

- 30-40% improvement in model accuracy
- 3-5x faster AI deployment
- 60-80% reduction in manual data prep
- Up to $3-5M annual savings through automation, faster time-to-value, and reduced compliance burden

Choosing the Right Partner

Scaling AI in healthcare demands deep domain expertise, cloud-native agility, and automated governance. IDMC, powered by CLAIRE AI, delivers these capabilities — automating ingestion, cleansing, mastering, and compliance for end-to-end data readiness.

Real World Impact

- Biopharma: 25% reduction in clinical trial cycle time, $4.8M annual ROI.
- Providers: 35% improvement in readmission prediction accuracy, $1.2M annual savings.
- Payer: $18M in fraudulent claims detected in year one.

The Future of Healthcare AI

Trusted data is the foundation for ethical, explainable, and scalable AI. With IDMC, healthcare leaders can transform fragmented data into governed, real-time intelligence — improving outcomes, strengthening compliance, and driving operational excellence. The question isn't whether to invest in trusted data — it's how fast you can begin.

About Informatica

Informatica's Intelligent Data Management Cloud (IDMC) is the world's leading cloud-native data management platform. Trusted by thousands of organizations across healthcare, financial services, retail, and other industries, IDMC enables organizations to discover, integrate, and govern data at scale—powering AI, analytics, and business intelligence. With CLAIRE AI at its core, IDMC automates data quality, master data management, data governance, and compliance—helping healthcare organizations build the foundation for responsible, scalable AI.

For more information, visit **www.informatica.com** or contact your Informatica representative.

Appendix: Resources

- Solution brief: **Unlocking Provider Innovation with Trusted Data and AI**
- Solution brief: **Unlocking Payer Innovation with Trusted Data and AI**
- Solution brief: **Unlocking Pharma Innovation with Trusted Data and AI**

About Us

About Informatica

Informatica from Salesforce is a leader in AI-powered enterprise cloud data management. Its Intelligent Data Management Cloud (IDMC) platform enables organizations to connect, manage and unify AI-ready data across the enterprise. With capabilities spanning data cataloging, integration, governance, quality, privacy, metadata management and master data management, Informatica supports a broad partner ecosystem and helps customers unlock the full value of their data and AI initiatives.

About Salesforce

Salesforce is the #1 AI CRM, empowering companies to connect with their customers in a whole new way through the power of artificial intelligence, data, and trust. For more information about Salesforce (NYSE: CRM), visit: **www.salesforce.com.**

Worldwide Headquarters
2100 Seaport Blvd.
Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
x.com/Informatica

CONTACT US

Where data & AI come to LIFE™

Informatica

IN09-5287-1225