RegenMed

# WHITE PAPER

## THE DEMISE OF HEALTHCARE DATA BROKERAGE

### ORTHOPEDIC REGISTRIES AS AN EXAMPLE

**August 11, 2026**

---

## TABLE OF CONTENTS

# INTRODUCTION

## Healthcare Is Broken Because Healthcare Data Is Broken

The United States spends more per capita on healthcare than any other nation, consuming over 18% of gross domestic product—nearly $15,000 per person annually. Despite this expenditure, which is roughly two and a half times the OECD average, U.S. outcomes across baseline health metrics consistently rank near the bottom among developed peers. Expenditures continue to rise without a corresponding improvement in health.

This macroeconomic failure is fundamentally an architectural data failure. The prevailing mechanism for clinical intelligence relies on centralized data brokers and registries that extract and silo medical information. This legacy model severs data from its clinical context, commoditizing raw records while failing to return actionable, verifiable evidence to the providers actively managing patient care.

Correcting this trajectory requires the structural dismantling of third-party healthcare data brokerage. The necessary evolution relies on transitioning from centralized data extraction to federated data architectures. By instituting sovereign data ownership and enforcing cryptographic data provenance, healthcare systems can replace delayed, retrospective reporting with secure, real-time clinical intelligence at the point of care.

## The Rise and Fall Of Healthcare Data Brokerage

### The Digitization of Medicine

The initial digitization of medicine was driven not by a clinical imperative to improve patient outcomes, but by regulatory mandates and the necessity to manage complex reimbursement models. Over the past two decades, healthcare providers were compelled to adopt a sprawling architecture of electronic medical records (EMRs) alongside independent revenue cycle management (RCM) and other platforms. These disparate systems were engineered to process an ever-expanding volume of medical billing codes and standardized ontologies rather than to optimize clinical care.

This digital expansion extended into research and population health, mandating independent software platforms to manage clinical trials, track social determinants of health

(SDOH), and implement value-based care (VBC) reporting. The resulting fragmented infrastructure imposes massive financial and administrative burdens, regularly consuming a significant percentage of hospital revenue while forcing health systems to fund multiple, non-interoperable data lakes.

## The Monetization of Personal Health Information

Virtually all clinical data originates as personal health information. The widespread adoption of digital medical records enabled a lucrative secondary market where third-party data brokers acquire heterogeneous patient records in bulk. These intermediaries aggregate, standardize, and scrub raw clinical data, selling the compiled datasets to life science companies, medical device manufacturers, and insurers for millions of dollars.

In this framework, patients sign broad consent forms during routine care without fully understanding how their data will be transferred or commercialized. Crucially, neither the patients nor the clinical providers who generate the data receive financial compensation. Furthermore, storing massive repositories of health data within centralized brokers creates profound privacy and security liabilities. Ostensibly de-identified datasets remain highly vulnerable to modern re-identification techniques, and centralized data lakes expose sensitive health information to persistent cyber threats.

## The Instructive Analogy of Financial Data

The infrastructure utilized to collect and commercialize medical data is undergoing a rapid, permanent dismantling. This transition is not following the gradual, decade-long adoption curves typical of past healthcare IT rollouts; rather, it is being forced by legal and commercial market pressures that will penalize organizations slow to adapt.

One strong historical parallel is the digitization of global financial markets in the late 1990s. For decades, financial trading was controlled by floor brokers and centralized exchanges that maintained monopolies on pricing information. They dictated transaction speeds, controlled access, and extracted correspondingly large fees. The advent of electronic trading networks removed these centralized intermediaries entirely. Buyers and sellers connected directly through distributed digital networks, accelerating data transfer from minutes to milliseconds. Floor brokers were rendered obsolete almost overnight, resulting in an immediate reallocation of wealth and control back to the market participants who

originated and consumed the data. Healthcare data is now crossing this same threshold.

### The Tectonic Shifts Already Underway

The legacy data broker model is actively collapsing under the weight of commercial realities, regulatory constraints, and advanced data architectures. Healthcare providers recognize that their clinical outcomes constitute highly valuable intellectual property. With the global healthcare data monetization market valued at $2.0 billion in 2025 and projected to reach $3.7 billion by 2033, providers are no longer willing to serve as uncompensated data generators for third-party intermediaries.

Simultaneously, new regulatory frameworks are exposing bulk data transfer to severe legal liability. Comprehensive privacy legislation, such as the Washington My Health My Data Act which took effect in March 2024, establishes strict consumer consent mandates for health data outside of HIPAA. These laws heavily restrict how brokers can collect and share derivative health inferences without explicit authorization, fundamentally threatening the traditional data brokerage model.

In response to these cyber and legal vulnerabilities, providers are moving away from centralized data lakes toward federated, decentralized networks. Rather than relying on the derivative billing data utilized by traditional brokers, modern architectures prioritize clinical ground truth. Utilizing tokenization and automated FHIR-to-OMOP mapping pipelines, hospitals and physician groups can now maintain physical custody, legal ownership, and direct commercial control over their data within closed-system collections. This shift successfully mitigates the vulnerabilities of centralized systems and effectively renders legacy data brokers obsolete.

## ORTHOPEDIC REGISTRIES: THE "CANARY IN THE COAL MINE"

### The Promise

Two decades ago, as joint replacement surgeries became increasingly common, the orthopedic community recognized a critical gap in patient safety. Clinical trials evaluated new implants on small patient cohorts for short durations, but no centralized system existed to track how these devices performed inside millions of individuals over the course of ten

or twenty years. Orthopedic registries were established to address this deficiency. Their foundational intent was to build a comprehensive, population-level database capable of monitoring long-term device survivorship across the healthcare continuum.

The theoretical utility of these platforms was immense. By aggregating procedural data from thousands of hospitals, registries were designed to function as a highly sensitive post-market surveillance system. If a newly released hip or knee implant began failing at abnormal rates, the registry would detect the statistical anomaly. This capability would allow medical societies and regulators to issue targeted recalls, isolate early clinical outliers, and remove defective hardware from the market.

Beyond surveillance, registries promised to elevate the standard of care by providing surgeons with direct access to massive datasets. This access was intended to determine exactly which surgical techniques and implant models produced the best outcomes for specific patient demographics. Had this potential been realized, registries would have established a universal, data-driven benchmark for orthopedic quality, permanently aligning daily surgical decisions with verified, long-term patient success.

## The Failings

After two decades, the structural limitations of the traditional registry model have become evident. The system has failed to adapt to modern clinical, economic, and technological demands, resulting in at least three distinct points of failure.

### Mismatched Value Exchange

Traditional registries require healthcare providers to bear heavy administrative and financial burdens. Hospitals and ambulatory surgery centers must pay subscription fees, fund expensive third-party IT integration costs to extract the data, and dedicate clinical staff hours to manual data submission. In exchange for supplying this foundational raw material, providers receive static, delayed reports that offer negligible actionable intelligence for daily clinical practice.

Meanwhile, registries retain the intellectual property rights and all commercial licensing revenue. This structural inequity creates a fundamentally unbalanced transaction where providers fund the centralized infrastructure and perform the labor, but capture none of the resulting economic or scientific value.

### Quantity Over Quality

Legacy registries prioritize scale, frequently citing databases containing millions of surgical records. However, this massive volume masks a severe lack of clinical utility. The data is aggregated through retrospective EMR extraction and delayed administrative billing codes. This methodology strips away granular clinical context, systematically omitting critical operative variables such as specific robotic alignment parameters or localized pain management protocols.

Accumulating millions of unstandardized, high-level records creates noisy datasets. These repositories demonstrate weak historical associations rather than the precise, actionable clinical causation required to improve patient outcomes or satisfy strict regulatory standards.

### The High Cost Of Institutional Inertia

Registries suffer from deep institutional inertia that renders them fundamentally incompatible with the 21st-century integration of artificial intelligence. Governed by slow-moving boards and legacy mandates, they cannot adapt to rapid technological or clinical shifts. They remain dependent on centralized infrastructure that necessitates manual data entry or expensive, localized software vendor integrations.

This dynamic reflects a classic "innovator's dilemma". Registries are too heavily invested in their legacy operations and existing industry sponsorships to dismantle their own infrastructure. Adopting modern, decentralized data architectures or AI-driven analytics would directly cannibalize their established value proposition and revenue streams. Consequently, they remain locked into expensive, delayed processes while the broader healthcare data market accelerates around them.

## Extrapolation To Other Healthcare Data Brokers

While orthopedic registries represent a well-intentioned, non-profit model initiated with active clinical participation, their architectural failures are magnified when the same centralized extraction model is deployed by commercial data brokers. The broader healthcare data market is increasingly dominated by for-profit aggregators accountable solely to public shareholders and private equity sponsors.

Unlike medical societies, these commercial entities possess no inherent mandate to improve

patient outcomes or elevate the standard of care. Their financial models rely on capturing medical records—often utilizing opaque data-sharing agreements and broad consumer consent mechanisms—to construct proprietary, centralized data lakes. The resulting commercial datasets are frequently expensive, narrow in their clinical utility, and fundamentally non-transparent regarding data provenance.

Because these for-profit brokers operate completely detached from the clinical point of care, they exacerbate the structural inequities of the legacy data model. Healthcare systems continue to function as uncompensated generators of highly valuable clinical intellectual property, while private equity-backed intermediaries extract billions of dollars in enterprise value without returning actionable intelligence to providers. This aggressive, non-transparent commercialization further validates the urgent necessity for healthcare systems to adopt closed-system data collection models, retaining strict sovereign control over their clinical assets.

## THE NEW HEALTHCARE DATA MODEL

### AI

The transition away from legacy data brokers is no longer theoretical; it is actively occurring. Artificial intelligence and modern, federated data architectures have fundamentally altered how clinical intelligence is gathered, analyzed, and deployed. Large Language Models (LLMs) now possess the computational capacity to ingest and analyze virtually all peer-reviewed medical literature, reasoning through complex clinical, scientific, and operational questions instantaneously.

Consequently, clinical AI adoption has accelerated rapidly. Rather than relying on delayed observational reports from centralized registries, healthcare providers now utilize AI platforms for real-time clinical education, diagnostic support, and treatment evaluation. Platforms such as OpenEvidence have achieved massive penetration—currently utilized by over 40% of U.S. physicians daily—functioning as high-speed literature lookup tools grounded directly in partnerships with organizations like JAMA and NEJM.

Furthermore, advanced AI applications can now securely access electronic medical records and protected health information (PHI) via digitized patient consent frameworks. This

frictionless access circumvents the manual abstraction processes that historically bottlenecked clinical research and commercial data extraction.

## Verifiable Evidence in Seconds, Not Years

The integration of AI into structured data networks establishes a new baseline for speed, resulting in strong market expectations for radically reduced costs across the entire healthcare ecosystem. This velocity is impacting every administrative layer and structurally forcing the dismantling of slow legacy processes.

For example, recent mandates from the Centers for Medicare & Medicaid Services (CMS)—specifically the Interoperability and Prior Authorization Final Rule (CMS-0057-F), which implements sweeping operational requirements in 2026—are driving the digitized, automated analysis of prior authorization requests by payers.

The requirement to share structured prior authorization data and clinical elements via real-time FHIR APIs forces payers and providers into near-instantaneous data exchange. Rather than waiting years for a medical society or third-party broker to aggregate data and publish a retrospective report, payers, providers, and regulators now possess the technological capacity to query structured networks and verify clinical evidence in seconds.

## Personal Health Record Security

The legacy data brokerage model relies on the extraction and centralization of protected health information (PHI) into massive, third-party data lakes. This architecture creates an indefensible security posture. While brokers claim to utilize strict de-identification protocols, the aggregation of longitudinal clinical records makes these datasets highly susceptible to modern computational re-identification techniques. Furthermore, centralizing millions of health records into single commercial repositories creates high-value targets for persistent, sophisticated cyber threats.

Federated data architectures invert this vulnerability. By utilizing closed-system data collections and cryptographic data provenance, clinical data remains physically and legally secured behind the originating healthcare provider's firewall. Intelligence and verified clinical algorithms travel to the data, rather than the data being extracted and exposed in a vulnerable, centralized third-party environment.

## Data and Evidence Ownership, Control and Monetization

The traditional centralized data broker is being removed from the transaction entirely. Historically, commercial brokers held a monopoly on aggregated clinical data because they were the only entities capable of financing the extraction of records from fragmented hospital systems.

Today, consumer-facing AI platforms can bypass institutional firewalls by connecting directly with patients. Utilizing digitized, HIPAA-qualifying authorizations, these platforms obtain direct consent to ingest comprehensive PHI. Crucially, patients are actively granting these platforms access to their complete medical histories at no cost, trading their data for the convenience of personalized AI health insights.

When this direct-to-consumer consent model is combined with an AI's capacity to instantaneously ingest and synthesize all publicly available peer-reviewed medical literature, the implications for traditional data brokerage are profound. The intermediary is completely disintermediated. AI platforms can generate superior, context-rich clinical intelligence directly from the patient source, effectively reducing the commercial value of third-party, bulk-purchased broker data to zero.

This structural shift restores fundamental data control to the creators—patients and physicians—allowing them to actively govern, protect, and monetize their clinical intellectual property without predatory intermediaries.

## Data Quality, Not Quantity

As the underlying data architecture transitions, the broader healthcare market has shifted its valuation from raw data volume to verified data quality. Historically, legacy registries justified their utility by touting databases containing millions of surgical records, despite the data lacking clinical granularity.

In the modern data economy, life science companies and medical device manufacturers require high-fidelity, structured clinical data to construct synthetic control arms and satisfy strict regulatory requirements. Precise, verifiable data points—such as localized pain management protocols or specific implant alignment angles—directly replace the need to recruit and fund expensive, traditional clinical trial cohorts. Consequently, massive volumes of stripped, unstandardized EMR data hold little commercial or scientific value compared to

smaller, highly structured, and cryptographically verified datasets.

### Humans "In The Loop"

However, delegating clinical analytics entirely to artificial intelligence introduces profound systemic risks. While AI accelerates data processing and democratizes access to medical literature, it possesses no inherent clinical judgment. Large Language Models remain susceptible to hallucinations and contextual misinterpretations, particularly when parsing complex, unstructured surgical notes.

To meet strict legal and regulatory standards—and to avoid severe medical liability—qualified humans must remain in the loop at critical junctures. Physician-validated, structured data networks ensure that the foundational information feeding these algorithms is accurate, safe, and clinically sound. AI is the engine of the new healthcare data model, but physician oversight remains the necessary cryptographic and clinical anchor.

## CONCLUSION

The United States healthcare system can no longer afford the administrative friction, financial drain, and clinical opacity imposed by legacy data brokers and centralized clinical registries. The technological tools necessary to dismantle this fundamentally unequal system—artificial intelligence, federated data architectures, and real-time structured data networks—are already operational and actively reordering the market.

By enforcing sovereign data ownership and demanding verifiable clinical quality over unstandardized quantity, healthcare providers can reclaim their intellectual property. The demise of healthcare data registries is not a future projection; it is a current reality. The institutions that adapt to this decentralized, AI-driven model will dictate the future standard of care, while those clinging to the centralized brokerage models of the past will be rapidly rendered obsolete.