zensar

HealthBridge: Governed Clinical Data Transformation on Databricks

A reusable approach to converting legacy healthcare feeds into FHIR-aligned data for quality, analytics, and AI.

A reusable pattern for making fragmented clinical data governed, usable, and ready for care, quality, analytics, and AI.

White paper

An RPG Company

Executive summary

Healthcare organizations do not struggle with a single interoperability problem. They face a recurring issue in clinical data transformation. Every acquisition, legacy feed, specialty system, EHR variation, and reporting initiative can introduce another source that must be understood, mapped, validated, governed, and made usable downstream.

HealthBridge converts HL7 v2, C-CDA, JSON, and FHIR STU3 into governed FHIR R4 within the customer's Databricks environment. The value is not another one-off interface. It is a reusable operating model: a single engine, reusable templates, configuration, validation evidence, and governed output that can be applied to the next similar feed rather than building from scratch.

The result is practical and measurable: fewer weeks per interface, less repeated build effort, faster validation, and faster access to governed clinical data for care, quality, analytics, and AI teams. The strongest near-term case is cost and capacity release; the revenue and risk cases are real but should be framed with discipline and validated through a pilot.

For Databricks teams, the value is customer adoption with purpose: health systems use the platform to solve real interoperability problems, keep clinical data governed, and reuse the same foundation for similar data-integration needs.

01. What is HealthBridge?

HealthBridge is a Databricks-native clinical data transformation framework. It converts HL7 v2, C-CDA, JSON, and FHIR STU3 into governed FHIR R4 inside the customer's Databricks environment. The value is not another one-off interface.
HealthBridge separates engine code, mapping templates, configuration, validation, and governed output, so similar feeds can be onboarded faster without having to start from scratch each time.

Zensar owns the healthcare mapping, implementation, validation, and production-readiness layer. Databricks provides the governed data foundation on which the framework runs.

At a glance

Who	Health systems with fragmented clinical data, including acquired hospitals, mixed EHR estates, specialty systems, legacy feeds, and teams responsible for quality, population health, analytics, AI, and governance.
What	A configuration-driven engine that converts HL7 v2, C-CDA, JSON, and FHIR STU3 into governed FHIR R4, and runs natively on Databricks.
When	When a health system needs to make clinical feeds usable quickly: acquisition integration, multi-EHR consolidation, specialty-system onboarding, FHIR modernization, quality reporting, population health, analytics, or AI readiness.
Where	Within the customer's own Databricks environment, using the governance, catalog, schema, and access-control model the health system already uses.
Why	Hand-built, interface-by-interface mapping cannot keep pace with the number of clinical data sources that health systems must integrate. Each new source that is not reusable adds time, cost, and risk to care, quality, analytics, and reporting work.
How	By separating code, mapping templates, and configuration, so a new source becomes a new template rather than a new project, the same engine is reused, not rebuilt.

02. The challenge: Recurring clinical data transformation

The visible symptom is a backlog. The root cause is reuse failure. Health systems repeatedly need to onboard, validate, govern, and reuse clinical feeds from acquisitions, EHR variations, specialty systems, legacy platforms, APIs, and reporting initiatives.

03. The root cause: Interface-by-interface delivery

The visible symptom is an excessive number of interfaces and insufficient integration capacity. The root cause is an operating model that treats each new source as a separate project.

HL7 v2, C-CDA, JSON, and FHIR STU3 often describe the same clinical reality in different structures. When each feed is mapped manually, the work is slow, hard to audit, and difficult to reuse.

Before and after: The operating model shift

Before: Interface-by-interface delivery	After: Reusable governed onboarding
Each new feed starts as a separate engineering effort.	Each new feed starts with reusable engine logic, templates, and configuration.
Mapping knowledge is trapped in individual projects and people.	Mapping lessons accumulate into reusable patterns that can help the next feed.
Validation often happens late and can require manual field-by-field checking.	Side-by-side source-to-output review makes validation easier to inspect.
Backlog grows whenever a new hospital, vendor, or message dialect appears.	Backlog can be reduced over time when reuse, governance, and access conditions are in place.
Downstream analytics and quality teams wait for integration work to complete.	Governed FHIR R4 output gives downstream teams a stronger clinical data foundation.

04. Where it sits in the value chain

HealthBridge operates at the unification and governance layer of the healthcare data value chain. It sits between fragmented source systems and downstream consumption by care, quality, analytics, and AI teams.

The leverage is clear: if unification is incomplete, every downstream process inherits that gap. Improving this layer gives the organization a stronger foundation for care visibility, quality reporting, population health, analytics, and AI.

Before and after: The operating model shift

Interoperability capability	Operational improvement	Business KPI it can support
Reusable source-to-FHIR conversion	Reduced repetitive interface-by-interface mapping.	Shorter feed onboarding time, subject to source access, quality, and approvals.
Governed clinical pipelines	Clinical data lands in a controlled, traceable environment rather than in a disconnected staging process.	Reduced integration friction and improved visibility into what has been converted and validated.
FHIR R4 standardization	Legacy feeds become available in a consistent target model for downstream consumers.	A larger share of the clinical estate becomes available for care, quality, analytics, and AI workflows.
Side-by-side validation	The review focuses on source-to-output evidence rather than manual field checking after the fact.	Faster validation cycles and clearer audit support, without implying any compliance guarantees.
Template-driven reuse	New feeds can reuse mapping patterns and delivery lessons from previous feeds.	Lower the interface backlog over time as reuse, governance, and access conditions are in place.

05. Use cases

Business problem	Where HealthBridge can help	How to frame it publicly
M&A clinical integration	Normalize acquired-site feeds into governed FHIR R4 so the enterprise can include those patients in downstream care, quality, analytics, and reporting workflows.	High-urgency example and strong first proof point.
Multi-EHR consolidation	Normalize feeds from Epic, Cerner, Meditech, ambulatory platforms, and other clinical systems into a single, governed, FHIR-aligned foundation.	Relevant for health systems operating mixed EHR estates after growth, acquisition, or partial standardization.
FHIR modernization	Convert legacy formats and older FHIR resources to a governed FHIR R4 foundation.	Core capability that requires older formats and newer standards to coexist.
Specialty-system integration	Bring lab, imaging, pharmacy, cardiology, oncology, and other specialty-system data into the governed clinical data foundation.	Useful when important clinical data resides outside the core EHR but still affects care, quality, and analytics.
Quality reporting and population health	Improve the completeness and traceability of clinical data that feeds quality measures, HEDIS, and population health analyses.	Supports better inputs but does not guarantee compliance or clinical outcomes.
Clinical data product creation	Use governed FHIR-aligned data to create reusable clinical datasets for care management, quality, analytics, and AI teams.	A downstream use once the source data has been validated and governed.
Analytics and AI readiness	Provide a cleaner, governed clinical data foundation for downstream analytics and AI use cases.	Position as data readiness, subject to source quality, governance, and validation.

06. How it works and why the design matters

Source systems emit HL7 v2, C-CDA, JSON, or FHIR STU3. HealthBridge applies the appropriate mapping template, converts the input into an FHIR R4 transaction bundle, and lands the output in governed Delta tables within Databricks.

The design advantage is separation of concerns: mapping lives in templates, environment settings live in configuration, and the engine is reused rather than copied. A new source becomes a new template and configuration path, not a new codebase.

Human verification remains explicit. A side-by-side source-to-output review speeds validation and makes it inspectable without removing clinical or governance oversight.

07. Defensible outcomes and KPIs

The strongest defensible value today is the combination of cost, effort, and speed. Revenue and risk benefits are real, but should be validated by pilot evidence. The figures below are modeled ranges, not guarantees.

Key metrics at a glance

50%-60% faster onboarding
Modeled reduction in elapsed feed onboarding time where conversion and mapping effort are the main bottlenecks.

50%-60% less build effort
Modeled reduction in engineering work per interface by reusing engine logic, templates, and configuration.

Validation in hours, not days
side-by-side source-to-output review can shorten validation cycles while keeping evidence inspectable.

These are modeled ranges, not guarantees. A pilot that uses the customer's own messages turns them into customer-specific numbers.

High confidence

Metric	What it means	Why we're confident
50%-60% faster onboarding	Elapsed time per feed, from request to production	Access provisioning and stakeholder sign-off are unchanged. This is the portion of the timeline the engine actually shortens.
50%-60% less build effort	Engineering hours per interface	Mapping is the automated piece; connection setup, testing coordination, and go-live are still real work.
Validation in hours, not days	Time to confirm converted output is correct	Side-by-side review replaces manual, field-by-field output checking

Directional, quantified after a pilot

Metric	Direction	Why we wait for real numbers
Rework from defects	Down	Verification catches issues before they move downstream, but the magnitude of the reduction depends on how error-prone the source feeds already are
Audit evidence effort	Down, not eliminated	Lineage and validation are built into the design rather than reconstructed after the fact, but "reduced" is honest; "eliminated" would not be
Interface throughput	Up	More feeds per engineer per quarter, if throughput (not access or governance approvals) is actually the bottleneck
Capacity	0.5 to 1.0 FTE-equivalent effort released across a multi-interface estate	This is capacity redeployed to higher-value integration and data work, not a headcount-reduction claim. The same team can clear more backlog when less effort is spent rebuilding similar mappings.

We also state the limits clearly. HealthBridge should be positioned as a pilot-ready framework, not as a completed customer case study. We do not claim universal conversion accuracy, clinical outcomes, compliance guarantees, or estate-wide cost of ownership. Accuracy depends on source data quality. Compliance remains the health system's responsibility. Cost depends on the size and condition of the clinical estate. The first completed pilot should establish measurable, customer-specific benchmarks and serve as the first formal proof point.

08. Who benefits

HealthBridge is most relevant when a health system has fragmented clinical sources, recurring onboarding needs, and an existing or planned governed data platform. M&A is one trigger; mixed EHR estates, specialty systems, FHIR modernization, and quality-reporting gaps are equally valid entry points.

Audience	What they get
The health system (the customer)	The potential to shorten onboarding of high-value clinical feeds, whether from acquired sites, legacy systems, specialty platforms, mixed EHR environments, or modernization efforts. The benefit is better visibility across integrated feeds and stronger support for quality, analytics, and operational decisions. Timing depends on access, source quality, security review, and operational approvals.
Databricks and platform stakeholders	Databricks benefits when customers use the platform for real clinical data problems: governed interoperability, reusable data integration, and trusted data for care, quality, analytics, and AI.
Implementation and engagement teams	A way to start with evidence: choose one painful feed, test it in the customer's environment, and decide what the results justify next.
The executive sponsor	A defensible case for the next step, based on measured results from the customer's own messages rather than generic claims.

The opening question is simple: which clinical data sources are still difficult to onboard, validate, govern, or reuse, and what does that delay cost the organization?

09. Why Zensar + Databricks

Open standards and public templates are the starting point, not the final service. Zensar adds healthcare mapping judgment, implementation discipline, Databricks-native packaging, validation evidence, governance alignment, and production support.

Databricks matters because HealthBridge runs in the areas where many healthcare organizations are already investing: governed data, Delta tables, Unity Catalog, analytics, and AI activation. Zensar makes the healthcare transformation workload real on that platform.

10. Objections we expect

Objection	How we answer it
Can we trust the conversion?	Trust here is earned by being inspectable, not by assertion. The side-by-side validation view shows source and output together, and the strongest way to answer this is to run a test against the customer's own messages rather than a demo dataset.
What about PHI and compliance?	HealthBridge is designed to keep data within the customer's governed Databricks environment throughout the pilot or deployment. Transformations are traceable, validation artifacts support review, and PHI handling remains governed by the health system's security, privacy, and compliance controls.
Why pay for something built on open source?	The open templates provide raw capabilities, not a production-ready solution. The real value lies in healthcare-specific mapping depth, rigorous engineering and delivery, Databricks-native packaging, validation evidence, governance alignment, and production support, turning publicly available templates into a system a health system can confidently deploy, operate, and rely on.
Will it fit our environment?	HealthBridge is not positioned as another middleware estate. It runs within the customer's Databricks environment, writes to governed Delta tables, and uses customer-controlled catalog and schema settings, enabling it to be evaluated as an extension of the existing data platform rather than as a separate system to procure, secure, monitor, and operate.

The pilot should demonstrate the engine's performance at the customer's selected volume and complexity. Open standards and permissively licensed templates reduce the risk of lock-in, but they do not eliminate the need for governance, validation, support, and operational controls.

11. Roadmap and expansion

HealthBridge starts with healthcare format conversion, FHIR bundle generation, resource flattening, and governed Databricks output. The first pilot may use a single interface, but the broader strategic question is whether the same framework can reduce repeated effort across a class of similar clinical transformation problems.

The expansion path includes additional clinical feeds, specialty systems, FHIR modernization, governed clinical data products, OMOP-oriented outputs, real-world evidence preparation, and analytics or AI readiness, all built on a shared FHIR-aligned foundation.

12. Pilot path

The pilot should start with a single high-value clinical feed in the customer's Databricks workspace. The goal is not to prove every use case at once. Instead, the goal is to measure whether reusable conversion can reduce onboarding time, build effort, validation effort, and downstream friction relative to the customer's own messages.

Pilot evidence scorecard

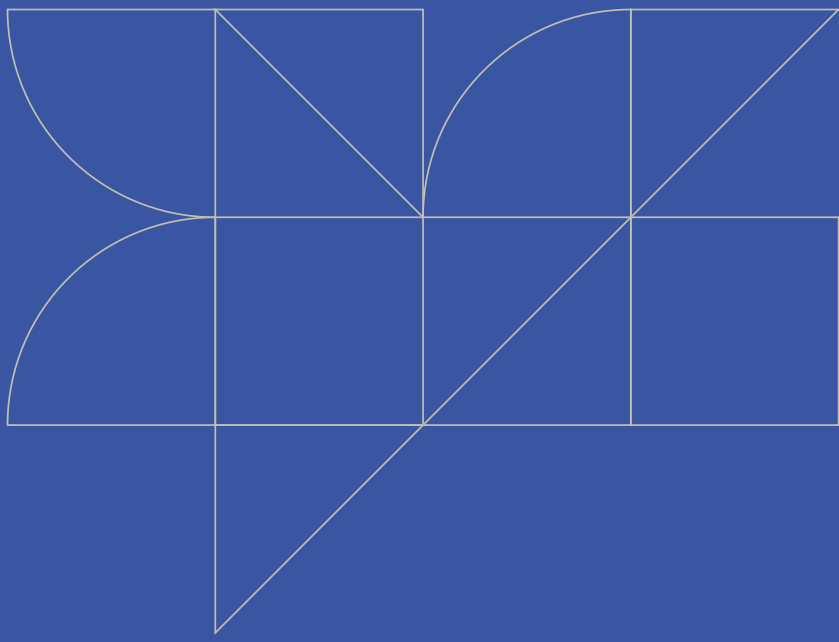
Evidence artifact	What the customer should inspect
Source message sample	The actual HL7 v2 or legacy payload selected for the pilot, with known source-system context.
Generated FHIR R4 output	The converted resource bundle and whether the output structure matches the intended target model.
Mapping path	The template and transformation route used from source field to FHIR output.
Validation result	Side-by-side review showing what passed, what required adjustment, and what remains source-quality dependent.
Measured time and effort	Elapsed time from source access to converted output, engineering effort required, and comparison to the current process.

The question before the pilot is simple: which business problem should be used to prove the model first? Should it be acquisition integration, specialty-system onboarding, quality reporting completeness, clinical data product creation, or analytics readiness?

The decision is whether to continue solving clinical data transformation one interface at a time or to run a pilot to test whether a governed, reusable approach can reduce effort across the next feed, the next source, and the next business problem.

Authored by

Daniel Gomez,
Practice Head - Data Engineering and Analytics
daniel.gomez@zensar.com



zensar
An RPG Company

At Zensar, we're 'experience-led everything.' We are committed to conceptualizing, designing, engineering, marketing, and managing digital solutions and experiences for over 145+ leading enterprises. Using our 3Es of experience, engineering, and engagement, we harness the power of technology, creativity, and insight to deliver impact.

Part of the $4.8 billion RPG Group, we are headquartered in Pune, India. Our 10,000+ employees work across 30+ locations worldwide, including Milpitas, Seattle, Princeton, Cape Town, London, Zurich, Singapore, and Mexico City.

For more information, please contact: **info@zensar.com | www.zensar.com**