b.well Connected Health: The $120M Infrastructure Play Quietly Powering Every Major Health AI Launch of 2026

MAR 30, 2026 • PAID

3 1 Share

Table of Contents

Introduction: The Un-AI Company

What b.well Actually Built

The Partnership Sequence That Changes the Story

The Technical Architecture Nobody Talks About

The Data Refinery as a Real Moat

Health Skills and the Product Roadmap Signal

The Competitive Risks Worth Taking Seriously

The Investment Thesis

Abstract

b.well Connected Health is a Baltimore-based health data infrastructure compar founded in 2015 by Kristen Valdes and Bryan Jones. Total disclosed funding is approximately $116M across 10+ rounds, with the most recent being a $40M Ser in February 2024 followed by $20M in Trinity Capital growth debt in July 2025. company has not publicly disclosed a valuation but is trading on Nasdaq Private Market.

Key facts: 2.4 million provider connections, 350+ health plan and lab connection FHIR-native canonical data model, 13-step proprietary Data Refinery, four SDK surfaces (Web TypeScript, Android Kotlin, iOS Swift in progress, AI via MCP), v label AI assistant called bailey, and a compliance product line tied to NCQA-cer digital quality measures.

Partnership sequence Oct 2025 through Mar 2026: Google (Oct 2025), SDK laun (Dec 2025), OpenAI/ChatGPT Health (Jan 2026), bailey white-label (Feb 2026), athenahealth point-of-care workflow (Feb 2026), Samsung/Kill the Clipboard (M 2026), Perplexity Health (Mar 2026).

Confirmed tech stack via sub-processor disclosure (effective Jan 5, 2026): AWS (infrastructure), Databricks (data processing), MongoDB (database), Redis (cachi Fivetran (ETL), Sigma Computing (analytics/BI), Groundcover (observability), Se (errors), Wiz (cloud security), Descope (identity), CLEAR (identity verification), CloudBees (feature flags), Mixpanel (app analytics), Iterable/Twilio (communicat Tonic.ai (synthetic data/de-identification).

Additional confirmed from GitHub and 2021 tech blog: Apache Kafka (event streaming), Elasticsearch (search), ClickHouse (OLAP analytics), GraphQL (quer layer), CQL (clinical quality logic), Kubernetes/Helm (deployment), Python (data engineering), Node.js (FHIR API server).

Open investment questions: current valuation unknown, no disclosed revenue or customer metrics, iOS SDK still listed as coming soon, Torch acquisition by Ope creates internalization risk, Microsoft chose HealthEx over b.well for Copilot He

Introduction: The Un-AI Company

There is a company sitting at the center of the most important AI health launche the past six months that has not raised a mega-round, has not been on the cover anything, and most people in health tech have never heard of. That company is b Connected Health, and the reason it does not get written about is exactly the rea

is worth understanding. It is not building AI. It is building the thing AI needs b
it can work in healthcare at all.

To get the full picture, back up to January 7, 2026. OpenAI announced ChatGPT
Health, its dedicated health AI product that allows users to connect their actual
medical records to their conversations with the model. The announcement was h
The coverage was everywhere. The thing most of that coverage missed, buried in
press release, was that the health data connectivity infrastructure powering the
thing was b.well. Not OpenAI's in-house data team. Not a bespoke API integrati
built by some well-funded startup. A Baltimore company with under $120M in
disclosed capital that had spent the better part of a decade building the plumbir
nobody else wanted to build.

That is not a lucky break. By the time the OpenAI deal landed, b.well had alread
signed Google (October 2025), launched the first SDK designed specifically for h
AI assistants (December 2025), introduced a white-label AI assistant called baile
(February 2026), partnered with athenahealth on bidirectional point-of-care data
sharing (February 2026), expanded a two-year Samsung partnership into a full K
Clipboard implementation at HIMSS (March 2026), and signed Perplexity for its
health product (March 2026). That is five major platform partnerships in five mo
each one with one of the most prominent technology companies on earth, all
funneling through the same data layer.

The framing that matters here is infrastructure versus application. Applications
acquired, disrupted, or commoditized. Infrastructure, if it becomes the standard
buried into the foundation and stays. The question for investors and founders
watching this space is whether b.well has actually become infrastructure, or whe
it just looks that way from the press release cadence.

What b.well Actually Built

The company's official description calls it a FHIR-native digital health platform
a connected health data network. That is technically accurate and almost comple
useless for understanding what it does. Here is the more useful version.

b.well spent roughly a decade quietly onboarding as a trusted third party to ever major payer and provider in the country, leveraging the information blocking ru and patient access API mandates created by the 21st Century Cures Act and CM interoperability regulations. By the time those regulations had teeth, b.well alre had two million provider connections and three hundred payer connections. Tha network is the foundation. Everything else, all the AI products, all the SDK surf all the enterprise software, runs on top of that foundation.

The network connects through multiple interoperability rails simultaneously. Pa Access APIs mandated under ONC (g)(10), which are the richest data pathway be they require USCDIv3 content including unstructured clinical notes. TEFCA Q for the national exchange layer, though TEFCA only operates at USCDIv1, mean is less data-rich than the direct API connections. Regional HIEs and HINs. CMS Button for Medicare. The VA. Proprietary pharmacy and lab networks. Payer cla APIs. The March 2026 technical blog post by Yelena Balin on the resource hub r the competitive argument plainly: companies that claim 90% coverage by counti EHR vendor logos are using a meaningless metric because a single physician can document care across four different EHR systems at four different organizations if you are only counting vendor relationships you are missing three of those four completeness requires NPI-level onboarding at individual clinic locations, not ju system-level agreements.

On top of the network sits the Data Refinery. b.well describes it as a 13-step proprietary process that has been in development for a decade. What that means practice is a pipeline that ingests data in every format healthcare has ever produ including X12 claims, HL7 v2 messages, C-CDA documents, CSV files, and JSO APIs, converts everything into standardized FHIR R4 resources, and then runs a sequence of cleansing, validation, deduplication, normalization, enrichment, and compression steps before the data touches any downstream application or AI sys The CTO Imran Qureshi published a detailed technical walkthrough of this in January 2026 on the resource hub, including a worked example of a single prescr generating six separate records across EMR, HIE, pharmacy, insurance, patient a and refill systems, each with overlapping but incomplete information, and how t

refinery reconciles those into one clean current-state record. That is not a marke story. That is the actual problem, and it is actually hard.

The refinery's commercial importance for AI is the 10x LLM token reduction cla Raw FHIR bundles are verbose, redundant, and expensive to process. A patient's complete medication history as raw FHIR JSON might cost several hundred toke per medication entry. The refinery compresses, reconciles, and structures that ir dense, AI-optimized representation. The compressed-fhir repository on GitHub technical implementation of that. Multiply the cost difference across every Chat Health user who connects their records and runs health conversations, and the economic case for b.well sitting between the patient's data and the language mod becomes obvious.

The Partnership Sequence That Change the Story

The most important thing to understand about the Q4 2025 through Q1 2026 partnership run is the pattern it reveals, not the individual deals.

Google signed in October 2025 with a focus on personalizing health data for Fit and AI use cases. That deal was initially vague in scope and the Google product had not yet shipped a dedicated health AI product. The b.well SDK for Health A launched in December 2025, described as the first SDK built specifically to pow healthcare AI assistants, with pre-built configurations for LLMs, chatbots, copil and agent-based systems. OpenAI signed in January 2026 and immediately used SDK to connect ChatGPT Health to real patient records. Samsung expanded its year partnership in March 2026 to implement Kill the Clipboard at HIMSS, turn Galaxy phones into portable health record portals powered by b.well's network a CLEAR's identity infrastructure. Perplexity announced in March 2026 the same it shipped Perplexity Health, with b.well handling the EHR connectivity layer wh Perplexity handled Apple Health and wearable data through Terra API.

Three things are notable about this sequence. First, every company that needed patient EHR data for an AI health product came to the same solution. That is no

coincidence. FHIR APIs have been public for years. These companies could have their own connectivity. They chose not to because the hard part is not speaking it is the decade of relationship-building with individual provider organizations, patient identity verification layer, the consent management infrastructure, and t data normalization work. None of those things can be bought off the shelf or replicated quickly.

Second, b.well specifically names Anthropic in its developer documentation alon OpenAI and Google as MCP client configuration targets. There is a dedicated configuration guide for Claude on the developer portal at developer.bwell.com. T is not a signed partnership announcement, but it is developer documentation tha does not get built unless there is either a live integration or an active conversatio about one.

Third, Microsoft went a different direction and used HealthEx for Copilot Healt which launched in March 2026. That is the most important competitive data poi the entire picture because it tells you the market is not winner-take-all and b.we not yet the universal standard. Understanding why Microsoft chose differently r both the limits of b.well's current position and the strategic risk on the table.

The Technical Architecture Nobody Talk About

The sub-processor disclosure from January 5, 2026 is the most unexpectedly reve document b.well has published. Sub-processor disclosures are required for healt companies under HIPAA and GDPR to inform clients of every vendor that may h access to their data. Most companies publish these as dry legal compliance documents. b.well's disclosure essentially maps the operational substrate of the platform.

Infrastructure is AWS. That is standard. The interesting part is that AWS Healthl Accelerator is also listed as a past investor, which means b.well has a strategic relationship with AWS that goes beyond vendor-customer. Databricks is confirm for data processing, meaning the heavy-lift transformation layer, including the

refinery pipelines, runs in Databricks on top of AWS. MongoDB is the database, which means the canonical FHIR documents and consumer-facing longitudinal records live in a document store rather than a relational database. Redis handles caching. Fivetran moves data for internal analytics. Sigma Computing is the inte BI and data visualization layer.

Beyond those core components, Tonic.ai is listed for de-identification and synth data generation. That is a specialized tool used by mature engineering organizat to create realistic but PHI-free test datasets that mirror production data structu tells you b.well has a sophisticated testing and QA environment that never expos real patient data to development workflows. CloudBees handles feature flags, m new product features ship behind flags that can be toggled per customer segmer user cohort. That is how you safely roll out AI features in a regulated healthcare environment without pushing untested functionality to all users simultaneously.

The 2021 technology team blog by Qureshi adds the components that predate th current sub-processor list and remain part of the stack. Elasticsearch is named explicitly alongside Spark, AWS, Python, and Node.js as core platform technolog The FHIR server GitHub repository confirms Kafka event streaming. The FHIR repo also has a clickhouse-init folder, confirming ClickHouse as an analytics lay The combination makes sense: MongoDB holds the canonical records and serves time application queries, Elasticsearch indexes those records for full-text search clinical note retrieval, ClickHouse handles population health analytics and HED measure calculations, and Kafka is the event bus that keeps all three in sync whe data arrives.

The GraphQL layer on top of the FHIR server, also named in the 2021 blog and confirmed by the repo, is a meaningful architectural detail for the AI use case. Standard FHIR REST calls return entire resource bundles even when you only n few fields. GraphQL lets the MCP tools request precisely the data needed for a g query, “give me active medications, most recent HbA1c, and current allergies for patient,” in a single request rather than multiple REST calls. Combined with the compressed-fhir library that reduces token footprint, this is the actual mechanis

behind the 10x LLM cost reduction claim. It is precision extraction plus compre not magic.

The platform name in the 2021 blog is Helix. The external branding has evolved internally the product still appears to be called Helix, visible in the Helm deploy repository at github.com/icanbwell/helm.helix-service. The FHIR server at fhir.perf.icanbwell.com runs on version 3.2.60, which provides a development environment URL for direct API testing without going through the enterprise p

The Data Refinery as a Real Moat

The refinery's public description includes the first few of its 13 steps: data conve (everything to FHIR), data cleansing (fixing typos, standardizing formats, correct invalid codes), and the beginning of a deduplication and quality assurance step. rest of the 13 steps are not publicly disclosed. That is deliberate intellectual pro protection. But the framing is enough to understand what it does at a category le

The problem the refinery solves is that FHIR compliance is largely fictional in th world. b.well's March 2026 technical blog by Balin points out that a substantial portion of data labeled as FHIR-compliant fails basic validation standards. Syste claim FHIR support while producing malformed or incomplete resources. Units inconsistent across systems. Patient names change and different systems capture different versions. Medications appear as multiple records representing transact history rather than current state. Clinical notes and care plans sit in unstructure document references that cannot be queried meaningfully without text extractio embedding. A standard FHIR aggregator that just pipes through what it receives passes all these problems downstream to the application layer. The refinery is wh intercepts and fixes them.

The competitive moat is not the refinery concept. It is the accumulated institutio knowledge of where healthcare data breaks. Every edge case that has been encountered and handled over ten years of onboarding providers, payers, and pharmacies is encoded in the refinery's 13 steps. A new entrant can build a FHIR aggregation layer in months. Building the clinical vocabulary mapping tables, th

deduplication logic that handles the same medication described as Lisinopril, lisinopril 10mg, and Zestril 10mg oral tablet in different source systems, the tem logic that resolves conflicting encounter dates from different systems, and the pa matching rules that handle name changes, address variations, and demographic c takes years and requires touching real production data across thousands of sour organizations.

The Tonic.ai sub-processor listing adds a layer to this story. b.well is not just ru the refinery on production data. It has a parallel synthetic data infrastructure th generates test datasets mimicking real production data patterns without exposin actual patient records. That capability is what allows the engineering team to bu and test new refinery steps against realistic edge cases without ever touching PH development. That is an engineering practice that compounds quality over time.

The oidc-auth-lib repository on GitHub is another signal along the same lines. b built and maintains its own OIDC authentication library on top of the Descope a CLEAR identity infrastructure. OIDC is the protocol layer; Descope handles ide orchestration; CLEAR provides biometric IAL2 verification. Three layers of ider infrastructure because the problem requires it. Consumer-mediated health data access requires that the entity authorizing data sharing is verifiably the patient, bad actor with a stolen password. IAL2 identity proofing means that the patient' identity has been verified against government documents, not just confirmed by email or phone number. Building that at scale and making it frictionless enough patients will actually use it is a hard problem that took years to solve.

Health Skills and the Product Roadmap Signal

The March 6, 2026 blog post from Qureshi titled Health AI Needs Skills, Not Ju Intelligence is the most forward-looking technical document on the resource hu deserves serious attention because it signals the next major product direction.

The argument starts with a worked example. A patient with rheumatoid arthriti Humira who has a documented egg allergy asks a simple question: do I need a flu

vaccine? A standard LLM answers correctly for a hypothetical healthy patient bu
dangerously incorrectly for this specific patient, because the right answer involv
knowing that Humira as a TNF inhibitor creates enough immunosuppression to
contraindicate live vaccines, that most flu vaccine formulations are egg-based, th
egg-free alternatives like Flublok exist, that timing vaccination mid-cycle betwe
biologic doses optimizes immune response, and that the specific health system's
rheumatology protocols, formulary preferences, and referral pathways matter fo
actually executing the recommendation. The LLM has academic knowledge from
medical training data. It does not have practical clinical knowledge from years o
treating immunocompromised patients. It does not have organizational knowled
about how a specific health system operates.

Health Skills are b.well's answer to this problem. They are described as structur
workflows written in plain English that teach bailey how to navigate specific clin
scenarios with practical and organizational context baked in. The flu vaccine ski
an immunocompromised patient is different from the flu vaccine skill for a healt
adult. Each skill encodes clinical judgment and health system-specific protocols
generic AI training data does not contain.

This is a significant product direction signal for three reasons. First, the Health
library becomes a proprietary asset that accumulates over time. Each skill deplo
represents institutional clinical knowledge that is not available to any AI platfor
without a b.well relationship. The more health systems deploy bailey with their
organizational protocols embedded in Health Skills, the wider the accuracy gap
between bailey and any competing AI assistant without that contextual groundir
Second, it positions bailey's Chat Completions API as something meaningfully
different from a thin FHIR retrieval wrapper over a foundation model. The Oper
compatible interface lowers the barrier to evaluation, but the actual value propo
is the combination of complete longitudinal records plus clinical workflow
intelligence plus organizational context. Third, Health Skills represent a new rev
line and a new switching cost. Once a health system's clinical protocols are enco
bailey's Health Skills library, replacing bailey requires re-encoding all of that
organizational knowledge in a new system.

The Competitive Risks Worth Taking Seriously

Four risks deserve serious treatment without the marketing filter.

OpenAI bought Torch in January 2026 for approximately $100M in an acqui-hire Torch was a four-person health records startup building technology specifically unify fragmented medical data for AI use. The team joined OpenAI's ChatGPT H service. That is a direct technical overlap with what b.well provides. OpenAI is currently a b.well customer. The Torch acquisition signals that OpenAI intends build at least some of that capability internally over time. The optimistic read is even with Torch's team, OpenAI will find the decade of network relationships ar refinery edge cases impossible to replicate quickly. The pessimistic read is that o OpenAI has its own capability, the b.well relationship becomes optional rather t essential.

Microsoft chose HealthEx over b.well for Copilot Health. Copilot Health launch March 2026 claiming connections to more than 50,000 US hospitals and provide organizations. That is a direct competing network claim. Whether HealthEx's coverage is comparable in data completeness to b.well's, or whether it achieves breadth at the cost of the depth the Balin coverage-versus-completeness argume describes, is not determinable from public evidence. But Microsoft's choice not b.well for its consumer AI health product is the single most important competiti data point available and the one that gets the least attention in coverage of b.wel recent partnership run.

Apple controls its own health data rails through HealthKit on iOS. The b.well iO SDK is listed as "coming soon" on the developer portal. Apple has no incentive t route patient data through b.well when it controls the native health data infrastructure for approximately 57% of US smartphone users. The Samsung partnership is Android-first by necessity, not preference. Perplexity Health conn to Apple Health through Apple's own integration and to EHR data through b.we That is a segmented architecture that reflects the reality of Apple's walled garde

a unified b.well-first strategy. Any investment thesis on b.well needs to account f
fact that the iOS path is structurally constrained.

The sub-processor list is missing Elasticsearch as a disclosed vendor even thoug
confirmed in the 2021 tech blog as a core platform technology. That discrepancy
either means Elasticsearch has been phased out in favor of ClickHouse or Datab
vector capabilities, or it means Elasticsearch is internal infrastructure that does
technically sub-process client data in the HIPAA sense. The distinction matters
because Elasticsearch's role in the full-text search and clinical note retrieval laye
the part most directly relevant to the AI SDK's claimed search capabilities. If b.v
migrating away from Elasticsearch toward a vector database or ClickHouse's nat
vector search, that is a meaningful architectural transition in progress.

The Investment Thesis

The core thesis is straightforward but requires resisting the temptation to map l
onto familiar digital health company archetypes.

b.well is not a patient engagement company. It is not an EHR. It is not a payer
technology company. The closest analogy is a data network utility, something mo
like Plaid for healthcare than like any of the companies it gets compared to in th
health tech press. Plaid built the bank connectivity network that every fintech
product depends on. Plaid's value was not the applications built on top of it. Pla
value was the network and the API abstraction layer. Visa eventually acquired it
$5.3B before the deal fell apart on antitrust grounds.

b.well's network, the 2.4M provider connections and 350+ payer connections wit
NPI-level depth and multi-pathway redundancy, took a decade to build and cann
replicated by a well-funded startup writing checks. That is the primary asset. Th
Data Refinery is the secondary asset, because it is the thing that makes the netw
output actually usable for AI rather than merely compliant. The Health Skills lib
and the consent/identity infrastructure are the emerging tertiary assets.

The strategic acquirer logic is strong on multiple dimensions. Google has an act partnership and is still building its consumer AI health product. Microsoft has a competitor product using HealthEx but may want b.well's network depth for fut builds. Amazon Web Services launched Amazon Connect Health in March 2026 competing agentic AI layer for clinical workflows and is already an investor and infrastructure provider. Samsung is a partner and investor. Any of the AI platfor currently using b.well's SDK could determine that internalizing the infrastructu cost is worth paying a strategic acquisition premium rather than continuing to p per-API-call licensing fees. The question of timing on any of those paths is genu unknowable.

The IPO path is murkier. b.well does not have the consumer brand recognition c Hinge Health or Omada Health, whose IPOs in 2025 reopened the digital health public market. It is infrastructure, and infrastructure companies tend to trade o revenue multiples that require disclosed revenue to anchor. There are no disclos customer metrics, no case studies with named outcomes data, and no public reve figures. That absence is either aggressive IP and contract protection (the enterpr contracts may prohibit disclosure), or it means the consumer-scale metrics are n large enough to publish without generating skepticism. A serious investor dilige process would probe that gap hard.

The honest summary is this. b.well has built something technically real and strategically positioned in a way that would have been hard to predict even three ago. The company bet on consumer-mediated, patient-controlled health data interoperability as the long-term infrastructure layer for the entire digital healtl ecosystem at a time when most of the industry was building closed, proprietary silos. That bet has paid off in the AI era because the AI companies cannot build they need without the infrastructure b.well spent a decade creating. The partner run of the past six months is validation, not origin. The risks are real, the compe dynamics are still evolving, and the Microsoft-HealthEx data point deserves mor scrutiny than it has gotten. But the core thesis, that owning the data normalizati and consent layer between fragmented healthcare records and AI systems is a du strategic position, holds.

3 Likes • 1 Restack

← Previous

Discussion about this post

Comments Restacks

Write a comment...

© 2026 Thoughts on Healthcare · Privacy • Terms • Collection notice
Substack is the home for great culture