

Royalty Pharma for Real World Data: The Case for a Royalty Financing Company That Buys Forward Rights to Health System Data Licensing Revenue and Turns Clinical Records Into a Securitizable Asset Class

JUL 15, 2026



Share



Part I Podcast free on Spotify.



Thoughts on Healthcare Markets & Technology



Part I: Royalty Pharma for Real World Data: A Royalty Financing Company That Buys Forward Rights to Health System Data Licensing Revenue & Turns Clinical Records Into a Securitizable Asset Class

Hospitals are sitting on a recurring, contractual, uncorrelated cash flow buried in 'other operating revenue.' Nobody finances it. Nobody prices it. That is the opportunity...

▶ Listen now

7 hours ago · Thoughts on Healthcare



Part II Podcast episode for paid subscribers only. Also available on Spotify.



Thoughts on Healthcare Markets & Technology



Part II: Royalty Pharma for Real World Data: A Royalty Financing Company That Buys Forward Rights to Health System Data Licensing Revenue &

Turns Clinical Records Into a Securitizable Asset Class

Hospitals are sitting on a recurring, contractual, uncorrelated cash flow buried in 'other operating revenue.' Nobody finances it. Nobody prices it. That is the opportunity...

▶ Listen now

7 hours ago · Thoughts on Healthcare

To listen to paid episodes in Apple or Spotify, link settings on those platforms (instructions inside the Podcast).

Table of Contents

- The observation: hospitals are sitting on miscellaneous revenue
- The template: what Royalty Pharma actually does
- The asset: what health system data licenses
- Why now: AI training demand, tokenization, and making money
- The model: forward purchase of data licensing royalties, structured three ways
- Underwriting: why the moat is a ratings problem, not a capital problem
- The portfolio math: diversification, duration, and what the yield has to be
- Securitization: the path from balance sheet to ABS
- The consent problem, the PR problem, and the Illinois problem
- Failure modes, honestly assessed
- The endgame: the reference pricer for clinical data

Thoughts on Healthcare Markets & Technology is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.

Abstract



Discover more from Thoughts on Healthcare Markets & Technology

Expert analysis of healthcare and life sciences markets, technology, investment, entrepreneurship policy, and AI — for investors, entrepreneurs,...

Over 4,000 subscribers

Enter your email...

Subscribe

By subscribing, you agree Substack's [Terms of Use](#), and acknowledge its [Information Collection Notice](#) and [Privacy Policy](#).

Already have an account? [Sign in](#)

- Health systems generate licensing revenue from de-identified clinical data sold into life sciences, registries, and increasingly AI training deals, but the revenue is lumpy, under-monetized, buried in “other operating revenue,” and invisible to lenders and rating agencies
- The proposed business buys forward rights to a share of that future licensing revenue in exchange for upfront capital, the same structure Royalty Pharma applied to drug royalties starting in the mid 1990s
- Hospitals are structurally capital starved, median margins remain thin post 2022, and traditional funding routes (muni debt, philanthropy, asset sales) are either tapped or dilutive of mission; a non-debt, non-dilutive advance against an asset the CFO barely knows exists is a genuinely new source
- The financier’s edge is underwriting: knowing which data assets (oncology depth, longitudinal completeness, linked claims, genomic overlays) command durable pricing requires data infrastructure expertise, not spreadsheet expertise, so the moat resembles a ratings agency more than a lender
- Pool royalties across dozens of systems and the portfolio becomes diversified enough to securitize, creating a new asset class the way drug royalties, music catalogs, and litigation claims each got financialized
- Honest risks: opaque pricing means soft marks, one state privacy law or one reidentification scandal impairs the whole book at once, revenue attribution through intermediaries is genuinely hard, and patients do not know they are generating collateral
- The long game is the same as the RADV essay’s punchline: whoever underwrites enough of these deals becomes the de facto ratings agency for clinical data quality, and reference pricers outearn lenders

The observation: hospitals are sitting on royalty streams and calling them miscellaneous revenue

Somewhere in the finance department of every large academic medical center there is a line item that nobody owns. It shows up in “other operating revenue,” it arrives in irregular chunks, and it comes from contracts the treasury team has never read: data licensing deals with life sciences companies, real world evidence vendors, registry operators, and lately, AI labs that need clinical text the way steel mills need coke. At a big AMC this line can run into the tens of millions annually. At a well organized multi state system with a dedicated data commercialization function it can run higher. At most systems it runs far below potential because nobody senior is paid to maximize it,

the deals are negotiated by whoever the pharma rep happened to email, and the CFO thinks about it the way one thinks about parking garage revenue, which is to say, never.

Now notice what kind of cash flow this is. It recurs, because pharma's appetite for real world data is structural and growing. It is contractual, with multi year terms and defined payment schedules. It is uncorrelated with the things that actually stress a hospital's P&L, meaning payer mix, labor costs, and census. And it is completely unfinanced. No lender advances against it, no rating agency models it, no investor prices it. It just sits there, an annuity wearing a trench coat, hoping nobody asks questions.

Every time capital markets discover a durable cash flow that the operating entity undervalues, the same thing happens: someone shows up with a check and buys the stream. It happened to drug royalties, music catalogs, franchise fees, litigation claims, GP stakes, and, in the truly degenerate cases, professional athletes' future earnings. It has not happened to health system data licensing revenue. This essay is about the company that makes it happen, and why the hard part is not the money.

The template: what Royalty Pharma actually figured out

Worth being precise about the analogy, because Royalty Pharma is one of the most misunderstood great businesses of the last thirty years. The lazy read is that they lend money to biotech. The accurate read is that they identified a cash flow, drug royalties owed to academic institutions, inventors, and small biotechs, that the holders systematically undervalued because the holders were not in the business of holding financial assets. A university tech transfer office with a two percent royalty on a blockbuster does not want twenty years of variable checks. It wants a building. Royalty Pharma bought the stream, gave them the building money, and pocketed the spread between what a diversified, duration matched, credit disciplined holder could pay and what an impatient, concentrated, unsophisticated holder would accept.

Three design choices made it work. First, they bought passive interests, not operations. No drug development risk, no commercialization responsibility, just a contractual claim on top line. Second, they diversified relentlessly, so that any single drug's patent cliff, safety recall, or competitive erosion was a portfolio event, not an extinction event. Third, and least appreciated, they built the best underwriting apparatus in the category. They could model a drug's revenue curve, generic entry timing, and indication expansion better than the sellers could, which meant they were

consistently the highest credible bidder while still buying below intrinsic value. The sellers were not fools. They were just in a different business.

Map that onto hospitals. A health system with a strong oncology data asset does not want to be in the business of optimizing licensing revenue over fifteen years. It wants to fund a tower crane, an Epic upgrade, or the nurse retention program that keeps the joint commission off its back. It will trade a slice of a cash flow it barely tracks for capital it desperately needs, at a discount rate that looks expensive to the hospital and cheap to anyone who can actually model the stream. The spread lives in the modeling gap, exactly where Royalty Pharma found it.

The asset: what health system data licensing revenue looks like today

Getting concrete about what is actually being bought. The market for real world data flowing out of provider organizations runs well into the billions annually and has compounded at double digit rates for a decade, driven by regulatory acceptance of real world evidence, the FDA's post 21st Century Cures posture, and pharma's realization that trial recruitment, label expansion, and HEOR arguments all run on linked longitudinal records. The buyers are pharma and biotech directly, the RWE analytics shops, CROs, registry operators, and now foundation model developers who have discovered that clinical notes are among the highest value text on earth and nearly none of it is on the open internet.

The supply side is fragmented in an almost comical way. A few systems run genuine data businesses with dedicated teams, standardized products, and per study or subscription pricing. A larger middle tier participates in federated research networks or aggregator relationships where an intermediary tokenizes, links, and resells, remitting a revenue share back to the system. And a long tail licenses episodically, one oncology dataset to one pharma sponsor, negotiated from scratch, priced by vibes. The pricing dispersion across these tiers for functionally similar data is enormous, which every reader who has sat on either side of one of these deals already knows and politely does not say at conferences.

The revenue characteristics matter for structuring. Contracts run one to five years. Some are flat subscriptions, some are per query or per cohort, some are milestone flavored when tied to specific studies. Counterparty credit is generally excellent, since the payers are large pharma and well funded analytics companies. Churn exists but the underlying demand driver, pharma's need for evidence, is about as secular as trends get. The stream most resembles a portfolio of specialty licensing royalties with

investment grade obligors and sub scale, unsophisticated licensors. Which is to say, it resembles the exact setup that royalty aggregators in other categories got rich on.

Why now: AI training demand, tokenization plumbing, and hospitals that need money

Three clocks all striking at once. The first is demand side and obvious: AI training deals introduced an entirely new buyer category with an entirely new willingness to pay. When model developers began signing licensing arrangements with publishers, forums, and stock photo libraries, the per token economics of high quality proprietary text got a public benchmark, and clinical documentation is scarcer, more structured, and more commercially consequential than almost any corpus that has traded. Systems that had never thought of their notes as an asset started fielding inbound. Several large health systems have already announced arrangements with AI developers around model building on their clinical data, and the ones that have not announced are mostly negotiating. This demand is young, poorly priced, and growing into a category that could plausibly rival traditional RWE licensing in size, which does interesting things to the value of a forward purchase struck against today's run rate.

The second clock is infrastructure. The reason this business was not buildable in 2015 is that nobody could meter the asset. Data left the building in bespoke extracts, attribution was a handshake, and auditing a revenue share meant trusting the intermediary's spreadsheet. The last decade of de-identification tooling, privacy preserving record linkage, and tokenization infrastructure changed that. Data can now be linked across sites and monetization can be tracked at the record and cohort level, which means a royalty owner can actually verify what was licensed, to whom, and for how much. Financialization always follows measurability. Music royalties securitized after digital distribution made streams countable. Same physics here.

The third clock is the buyer of capital. Hospital finances remain structurally ugly. Median operating margins recovered from the 2022 trough but sit at levels that do not fund the capital cycle, roughly a third of hospitals run negative, capex backlogs from the pandemic years remain unworked, and the muni market charges real spreads to anything with a shaky coverage ratio. Meanwhile the policy environment, site neutral payment proposals, 340B contraction, Medicaid funding pressure from the 2025 reconciliation law, all points toward more revenue stress, not less. A CFO staring at that picture will take a meeting about non-dilutive, non-debt capital secured by an asset the rating agencies do not even model. The pitch writes itself: this money does

not touch your debt covenants, does not show up in your leverage ratios, and comes from a line item your board has never asked about. Try getting that from a bond deal.

The model: forward purchase of data licensing royalties, structured three ways

The core transaction is a purchase, not a loan. The financier pays upfront cash for the right to a defined share of the system's data licensing revenue over a defined term. Structuring it as a true sale rather than secured lending matters for the hospital's balance sheet treatment, for bankruptcy remoteness, and frankly for the sales motion, since "we buy a royalty" opens doors that "we lend against your data" slams shut.

Three product variants, in ascending order of ambition. The first is the synthetic royalty on existing contracts: buy a percentage of revenue from the system's current book of licensing agreements, priced off contracted cash flows with modest renewal assumptions. Lowest risk, tightest pricing, essentially factoring with a longer tail, and the right wedge product because diligence is confined to reading contracts that already exist. The second is the whole asset royalty: buy a share of all data licensing revenue the system generates over, say, seven to twelve years, existing and future contracts alike. This is where the real value is, because it captures the AI training upside and the growth from professionalizing the system's commercialization function, and it is where underwriting gets genuinely hard, because the buyer is now pricing a growth curve rather than a contract stack. The third is the development deal: upfront capital plus an operating partner arrangement where the financier's team actually builds the system's data product capability, cleans the asset, stands up the licensing function, and takes a larger royalty for it. This is the Royalty Pharma to launch royalty progression, capital plus capability, and it is the version that turns the highest returns because the financier is manufacturing the very cash flow it owns.

Deal sizing sketches, held loosely. A large AMC doing twenty million a year in licensing revenue with visible growth might sell a quarter of the stream for ten years. At the discount rates this risk initially commands, mid teens unlevered, the check lands somewhere in the thirty to fifty million range, which is real money to a hospital and a rounding error against its bond stack. Do thirty of those across systems of varying size and the platform holds a royalty portfolio north of a billion in deployed capital within a few years, which happens to be roughly the scale at which securitization conversations start.

Underwriting: why the moat is a ratings problem, not a capital problem

Capital is a commodity here. Any credit fund can write the checks. What the credit funds cannot do is answer the only questions that matter: is this data actually worth anything, to whom, for how long, and at what price. That is not a finance question. It is a data infrastructure question, and it is brutal.

Consider what real diligence on a health system data asset involves. Depth and completeness by clinical domain, because a deep, molecularly annotated oncology cohort with linked outcomes prices at a multiple of general primary care records. Longitudinality, because pharma pays for patient journeys, not snapshots, and a system with high patient leakage produces fragmented journeys worth a fraction of a closed system's. Linkage quality, meaning whether the records can be tokenized and joined to claims, mortality, labs, and SDOH sources without match rates collapsing. Documentation quality and structured data discipline, since a system where half the meaningful clinical content lives in unstructured notes with idiosyncratic templates is sitting on an asset that costs more to refine than commodity data sells for, though ironically that same messy text is exactly what the AI training buyers want, which creates a genuinely funny valuation tension where the pharma buyer and the model developer are paying for opposite properties of the same corpus. Then contract diligence: what rights has the system already granted, exclusively or otherwise, to which intermediaries, with what revenue shares and what termination provisions, because the number of systems that have quietly encumbered their best assets in old aggregator agreements they no longer remember signing is not small.

Nobody can run that diligence without having operated in the plumbing. The founding team for this business is not ex Blackstone, it is people who have priced data licensing deals from inside the ecosystem, who know what a pharma evidence team actually pays for a linked oncology cohort versus what the intermediary reports back to the health system, and who can look at a system's Epic instance and estimate refinement cost within a defensible range. That knowledge is scarce, it does not live in banks, and it compounds with every deal underwritten, which is the whole point. The tenth deal is priced with proprietary comps from the first nine. By deal forty the platform has the only real transaction database in a market whose defining feature is that nobody knows what anything costs. In a market with total price opacity, the party with the comps is not a participant, it is the market.

The portfolio math: diversification, duration, and what the yield has to be

The honest actuarial picture. Individual royalties carry real idiosyncratic risk: a system gets acquired and the buyer has its own data strategy, a key aggregator

relationship terminates, an Epic migration torches two years of data continuity, a chief data officer quits and the licensing function reverts to voicemail. Portfolio construction has to assume some streams underperform badly, which pushes required gross returns on individual deals into the mid to high teens, with the whole asset and development variants priced above that. That sounds expensive until compared against the hospital's alternatives, which are issuing taxable debt at meaningful spreads, selling actual buildings, or begging the philanthropy office to find another naming opportunity. Against dilution of mission assets, a royalty on parking garage revenue's weird cousin is cheap.

Correlation is the more interesting risk, and it is two sided. The bad correlation is regulatory: a federal privacy statute with a data sale consent requirement, a state law wave following the genetic privacy litigation pattern, or a public reidentification scandal that makes systems radioactive about licensing, any of which hits every royalty in the book simultaneously. The good correlation is the AI demand shock, which lifts every royalty simultaneously and is, on current evidence, the stronger force. A portfolio buyer is net long the proposition that clinical data becomes more valuable over the next decade, hedged by contractual floors and diversification against the proposition that any single asset craters. Readers can decide for themselves which side of that trade they would rather hold, but the demand curve for clinical evidence has gone up and to the right through every regulatory regime since HITECH, and the AI bid arrived on top of it, not instead of it.

Duration matching is the quiet structural advantage. These royalties throw off cash over five to fifteen years, which is poison for a typical PE fund clock and perfect for insurance balance sheets, pensions, and permanent capital vehicles. The platform likely ends up structured the way the drug royalty and music royalty players ended up, as a permanent capital company or a listed vehicle, because the asset wants owners who never have to sell.

Securitization: the path from balance sheet to ABS

The end state is an asset class, and the road is well paved by prior financializations. The sequence is always the same: aggregate enough homogeneous cash flows, build a servicing and reporting standard, get a rating agency comfortable with the collateral analysis, and issue notes against the pool. Music catalogs did it, starting with the Bowie bonds in the nineties and maturing into a routine ABS category three decades later. Drug royalties did it. Litigation finance is mid transition. Whole business securitization taught agencies to rate franchise royalties from fast food chains, which

means the analytical machinery for rating weird royalty collateral exists and gets bored easily.

Data licensing royalties are actually good collateral by ABS standards: investment grade obligors on the underlying licenses, contractual cash flows, low operational intensity, and diversification across dozens of licensors and hundreds of license agreements. The hard part is standardized reporting, which is precisely what the tokenization and metering infrastructure now enables and what the platform's servicing function exists to produce. First securitization plausibly happens once the pool clears a billion in collateral value, notes price inside the platform's equity return by a wide margin, and the spread between mid teens asset yields and single digit note coupons becomes the engine that makes the equity math genuinely attractive. At that point the business is no longer a fund, it is an origination and servicing machine with a securitization takeout, which is the business model of every consumer lender worth owning, transplanted into the strangest collateral imaginable.

The consent problem, the PR problem, and the Illinois problem

Now the section that has to be written carefully, because the essay's whole thesis involves financializing an asset that patients generate without knowing it. The legal baseline is that properly de-identified data under the HIPAA standard is not protected health information and can be licensed without patient authorization. That baseline has held for twenty five years and the entire RWE industry stands on it. But the baseline is not the whole picture. State laws are moving, with Washington's My Health My Data act and its private right of action as the template that plaintiff lawyers would like to export everywhere, and the class action bar has spent recent years mining biometric and pixel tracking statutes for settlements against health systems, a pattern that suggests exactly how a data royalty portfolio would be attacked. The Illinois problem is shorthand for the scenario where one state's statute plus one aggressive litigation theory reprices consent expectations nationally, the way BIPA did for biometrics.

The reputational version is worse than the legal version. A single credible reidentification demonstration involving a marquee health system's licensed data would freeze the market for quarters, and the headline "Wall Street firm owns the rights to your hospital records" writes itself regardless of technical accuracy. The platform's response cannot be to hope. It has to build consent and transparency into the product as a feature, favoring systems with governance frameworks and patient communication practices that survive scrutiny, pricing governance quality into the

royalty the same way ESG got priced into everything else, occasionally sincerely. There is also a genuinely constructive version available: structures where a defined slice of licensing revenue funds patient facing benefits, charity care offsets, or community health programs, which converts the optics problem into a differentiator and gives the health system a public answer to the question of who benefits. Cynics will note this resembles reputation laundering with extra steps. Cynics are welcome to explain the alternative that survives a Sunday New York Times feature.

Failure modes, honestly assessed

The full stress test, without flinching. Marks are soft, because pricing opacity cuts both ways: the same comp database that gives the platform an underwriting edge also means its portfolio valuations are unverifiable by outsiders, which is how every private credit adjacent strategy eventually gets accused of marking fiction. The discipline is contractual cash flow coverage, valuing the book off collected dollars rather than modeled upside, and resisting the urge to write up the AI optionality before it converts to contracts. Attribution is hard, because revenue flows through intermediaries who net fees, bundle assets across systems, and report on lag, and a royalty owner without audit rights and metering visibility is trusting a spreadsheet, which is not a strategy. The fix is contractual audit rights plus the tokenization level tracking discussed earlier, and walking from deals where the intermediary stack cannot be seen through, however attractive the headline yield.

Concentration in disguise is a subtler one. Dozens of health system counterparties can still collapse into a handful of underlying demand nodes, a few large pharma evidence budgets and a few AI labs, and if two of those buyers merge or retrench the diversification was cosmetic. Obligor level exposure analysis has to look through the licensor to the licensee, which no hospital currently does and the platform must. Then the technology reversal risk: synthetic data gets good enough to substitute for real records in some use cases, or foundation models trained on the existing corpus stop needing incremental data, and the terminal value assumptions compress. Probably overrated as a near term threat, since evidence generation for regulators requires real patients and provenance, but a ten year royalty buyer has to hold some probability on it. And finally the boring one that kills more specialty finance companies than any exotic risk: deployment pressure. The moment the platform raises more capital than the origination pipeline supports, underwriting standards decay, the development deals get done with systems whose data should never have been commercialized, and the vintage that results poisons the securitization story for everyone. The graveyard of royalty aggregators is full of firms that died of success in fundraising.

The endgame: the reference pricer for clinical data

Same punchline as every great underwriting business, worth stating plainly. The royalty checks are the revenue. The comps are the company. After fifty transactions the platform holds the only standardized, diligence grade dataset on what clinical data assets are actually worth: which domains command premiums, which intermediary structures leak value, which governance postures survive legal weather, what AI training rights should cost relative to RWE licensing rights. That knowledge makes the platform the reference pricer in a market that currently has no prices, only anecdotes.

Reference pricers get paid from every direction, and the map is familiar from the ratings agencies and the title plants. Health systems pay for asset valuations before negotiating with aggregators. Pharma pays for benchmark data on what evidence should cost. Acquirers of health systems pay to diligence the data asset nobody else can value. Investors in the eventual ABS pay for surveillance. And the platform's own origination gets cheaper and better with every data point, which is the compounding loop that turns a clever financing idea into an institution. Hospitals spent a century learning that their real estate was an asset someone would finance. The records turned out to be worth more than the buildings. Someone is going to build the company that proves it, and the only real question is whether it gets built by people who understand the plumbing or by a credit fund that learns the hard way why match rates matter

Subscribe to Thoughts on Healthcare Markets & Technology

By Thoughts On Healthcare Markets & Technology · Hundreds of paid subscribers

Expert analysis of healthcare and life sciences markets, technology, investment, entrepreneurship, policy, and AI — for investors, entrepreneurs, hospital and insurance executives, and physicians navigating the business of healthcare.

By subscribing, you agree Substack's [Terms of Use](#), and acknowledge its [Information Collection Notice](#) and [Privacy Policy](#).



1 Like

← Previous

Discussion about this post

Comments

Restacks



Write a comment...