

Cureledger: The Data Layer for Rare Disease Drug Development

April 2026

Executive Summary

Rare disease has been the structural exception in drug development for four decades. Conventional pharmaceutical economics require each program to recover its cost against an addressable market large enough to justify the labor it consumes, and rare disease patient populations are by definition too small to clear that threshold. Approximately 300 million people live with one of more than 10,000 rare diseases worldwide, a population comparable in size to the United States. Fewer than one in ten of those conditions has an approved therapy. The unmet need is the largest in medicine, and the rare disease community has pursued cures for decades against economics that have not allowed the field to deliver them.

Those economics have changed. Agentic artificial intelligence has collapsed the per-hypothesis cost of drug discovery toward zero, and the Food and Drug Administration has opened a regulatory pathway for individualized therapies that do not fit the conventional randomized controlled trial model. The [Plausible Mechanism Framework draft guidance](#) issued February 23, 2026 accepts mechanism-anchored evidence as substantial evidence of effectiveness for genetic conditions with known biological cause, on the condition that high-quality natural history data support the mechanism. Taken together these two shifts make rare disease addressable in a way it has not been for forty years.

One constraint remains. Agentic research and the Plausible Mechanism Framework both require a data layer that is longitudinal, consented, reusable across programs, and provably compliant. The existing deidentified health data economy, dominated by aggregators including IQVIA, Datavant, Truveta, Komodo Health, and Tempus AI, does not supply such a layer for rare disease. The commercial aggregators serve common diseases where scale is achievable through back-end tokenization of records that patients never consented to monetize. Rare disease requires the opposite posture: direct, consented, patient-level longitudinal relationships that survive the exit of any individual sponsor and that compound in value as successive programs reuse the same evidence base. The consent failures exposed by the 23andMe bankruptcy and by the Lacks family's settlement with Thermo Fisher have made the inadequacy of the existing model visible to regulators, state attorneys general, and the public in ways that will not reverse.

Cureledger is the data layer the new regime requires. Patients contribute their health data into a Life Data Trust, a legal trust whose trustee administers the data under terms the patient sets and under a fiduciary duty the patient can enforce. The data is de-identified by default and subject to a conditional re-identification path that the patient controls and the trustee operates, used only to inform the patient of developments relevant to them. Revenue from authorized reuse flows back to contributors pro rata. Smart contracts and cryptographic proofs render the trustee's operations verifiable to contributors, regulators, and data purchasers without requiring any party to trust Cureledger itself. The architecture is purpose-built for rare disease and generalizes without modification to the broader drug development data economy, a market whose commercial ceiling is materially larger than the rare disease vertical alone.

Cureledger's thesis is that operational agentic AI, the Plausible Mechanism Framework, and a fiduciary-grade data layer are the three components of the next regime of drug development. Two are arriving under their own momentum. The third requires purpose-built infrastructure, and that infrastructure is what Cureledger delivers. Rare disease is the first proof. The larger market follows.

This document is organized in five parts. Part I establishes why the present moment is the moment for the data layer to exist. Part II describes how the existing data layer fails rare disease and why incremental repair falls short of what the new regime requires. Part III describes the Cureledger architecture. Part IV addresses the market and the commercial strategy. Part V addresses the competitive moat.

Part I. Why Now

1.1 The Labor Bottleneck Has Dissolved

For most of the modern pharmaceutical era, drug development has been rate-limited by the cost of human researcher hours. Every hypothesis advanced through a program required scientists to design, run, and interpret experiments, and the per-hypothesis cost

set a floor under which conditions were worth studying at all. Rare diseases sat below that floor. The total addressable market for a therapy targeting a small, geographically dispersed patient population could not recover research and development labor cost under conventional pharmaceutical economics, and programs that might otherwise have been scientifically tractable were not funded.

That cost structure has changed. Agentic artificial intelligence can now generate, rank, and stress-test clinical and biological hypotheses continuously and in parallel, across thousands of conditions, at a marginal cost that approaches zero. The constraint that once bound drug development to large-indication economics is dissolving. What remains binding is the quality of the substrate on which those agents reason. Hypothesis generation has become cheap; reliable data has become the scarce input.

The consequence for rare disease is structural. Under the old economics, small patient populations made rare disease programs uneconomic regardless of scientific tractability. Under the new economics, the binding constraint is whether reliable, longitudinal, reusable data about those patients exists in a form an agent can trust.

1.2 The Regulatory Pathway Has Opened

On February 23, 2026, the Food and Drug Administration issued draft guidance titled [Considerations for the Use of the Plausible Mechanism Framework to Develop Individualized Therapies That Target Specific Genetic Conditions With Known Biological Cause](#), Docket FDA-2026-D-1256. The draft accepts mechanism-anchored evidence as substantial evidence of effectiveness for highly targeted individualized therapies where the biological cause is known. The comment period closes April 27, 2026.

The Plausible Mechanism Framework represents a structural shift in how the agency evaluates effectiveness in rare disease. Randomized controlled trials remain the default in most regulatory contexts, yet the PMF recognizes that for many rare and ultra-rare genetic conditions a conventional trial is neither scientifically informative nor ethically obtainable. The draft guidance contemplates evidence generated across product variants, master protocols, and cross-program interoperability; it recommends early initiation of observational protocols; it relies explicitly on natural history data as a reference against which treatment effect is measured.

The PMF consolidates a line of precedent that has been building since 2019. Milasen, the bespoke antisense oligonucleotide developed for a single patient with a Batten-disease variant, established the modern n-of-1 regulatory framework when it was reported in the [New England Journal of Medicine in October 2019](#). The [n-Lorem Foundation](#) now provides free-for-life individualized antisense therapies to nano-rare patients under parallel arrangements. The FDA's [Rare Disease Innovation Hub](#), launched in 2024 and jointly operated by CBER and CDER, coordinates ultra-rare programs across centers. The PMF consolidates these threads into a general framework that other sponsors can invoke.

A regulatory pathway exists today that did not exist two years ago. The pathway assumes the existence of high-quality natural history data. The agency has not yet addressed where that data comes from, who holds it, how it is protected, or how it persists beyond the sponsors who generate it.

1.3 The Missing Piece is the Data Layer

Artificial intelligence agents are only as reliable as the substrate on which they reason. In rare disease, the substrate is broken along every dimension that matters to a regulator, a sponsor, or an agentic research system.

Rare disease data is fragmented across electronic health record systems that do not interoperate, across patient registries operated by disease-specific foundations with no shared data standard, and across clinical trial archives that belong to sponsors whose programs may outlive them and may not. It is de-identified under HIPAA Safe Harbor rules that, as Part II explains, do not deliver meaningful privacy protection for small populations. It is consented under terms written for a single study, terms that contemplate neither downstream reuse by another program nor secondary analysis by an autonomous agent. It is rarely longitudinal; most patients appear in a given dataset for the duration of a study and then disappear from it. It is held under custody rules that assume the sponsor persists, and in rare disease sponsor discontinuation is a predictable event rather than a corner case.

A reliable agentic research regime cannot be built on a substrate with these properties. Hypothesis generation that is cheap in compute remains expensive if every candidate hypothesis must be re-grounded in data whose provenance, completeness, and legal reusability are in doubt. The bottleneck moves upstream from researcher hours to data quality, and data quality in rare disease today is a problem of infrastructure, legal design, and governance.

The infrastructure problem is solvable. The regulatory and computational halves of the new regime are arriving under their own momentum. The data half is the piece that requires purpose-built work.

1.4 Rare Disease as the Proving Ground

Approximately 300 million people globally live with one of more than 10,000 rare diseases. Fewer than one in ten of those conditions has an approved therapy. The population is small relative to the diseases that dominate pharmaceutical portfolios and large in absolute terms, comparable to the population of the United States. It is geographically dispersed, medically heterogeneous, and served by a community of patient advocacy organizations, family foundations, and independent researchers that has historically driven the field's most important advances.

Rare disease families are the most motivated cohort in medicine. They bear the full weight of conditions for which treatment, when available at all, is typically lifelong and often unaffordable. They cannot rely on population-scale risk pooling to distribute the cost of therapy, because the populations are by definition too small. The unit cost of a rare disease therapy does not decline with time. Access can be lost when a sponsor exits a program or when a payer revises a formulary. These families do not have the luxury of treating lifelong therapy as a satisfactory outcome. Cures are the honest goal, and the regulatory and scientific framework is finally bending toward that goal.

The same properties that make rare disease the hardest case under the old drug-development economics make it the ideal proving ground under the new. Motivation among patients and families is high; consent friction is low; the value of a single reliable longitudinal record is multiple orders of magnitude greater than in a common disease, because the denominator of comparable patients is small. A data layer that works for rare disease will satisfy the harder case by construction. Carrying the same infrastructure into larger indications is the easier direction of generalization, and the commercial upside of doing so is the much larger market for verifiable multidimensional data across drug development generally.

The remainder of this document describes why the existing data layer cannot serve rare disease, what an adequate alternative looks like, and how the alternative generates value for contributors, sponsors, and investors in the same motion.

Part II. Why the Data Layer Breaks Today

The failure modes are cumulative. Any one of them would be sufficient to make rare disease research harder than it needs to be. Taken together, they make the current data layer unsuitable for the regulatory and computational regime that the field is about to enter.

2.1 Fragmentation and Sponsor Custody Failure

Rare disease data is generated in four principal settings: electronic health records maintained by providers, patient registries maintained by foundations and academic centers, clinical trial archives maintained by sponsors, and patient-kept records maintained by families. These settings do not share a data model, do not share consent terms, and in most cases do not share access. A researcher assembling a cohort for an ultra-rare condition commonly negotiates separately with each of these sources, re-consents the relevant patients where consent is obtainable at all, and reconciles divergent data dictionaries by hand.

Custody of the trial and registry data is the more fundamental problem. Under ICH Good Clinical Practice, sponsor retention obligations run only so long as the sponsor exists as a legal entity. No federal rule specifies the disposition of trial or natural history data when a sponsor becomes insolvent, is acquired, or quietly discontinues a program. In rare disease, sponsor discontinuation is predictable. The total addressable market economics that have always discouraged entry into rare disease also discourage persistence, and the longitudinal evidence base that the Plausible Mechanism Framework depends on is therefore held by exactly the parties least likely to preserve it across the lifetime of a patient cohort.

The evidence base survives only as long as its weakest custodian. Without a change in how custody is assigned, the data that a successor program would have relied on will not be there when the successor program arrives.

2.2 HIPAA Safe Harbor Does Not Protect Small Populations

The HIPAA Privacy Rule permits de-identification of protected health information either through the Safe Harbor method, which enumerates eighteen identifiers to remove, or through the Expert Determination method, which certifies statistically low risk of re-identification. Safe Harbor is the dominant method in practice because it is mechanical and auditable. It is also mathematically inadequate for small populations.

Rocher, Hendrickx, and de Montjoye showed in [Nature Communications 10:3069 \(2019\)](#) that, using generative models trained on demographic and health attributes, 99.98 percent of Americans can be re-identified in any reasonably rich dataset from as few as fifteen attributes. The paper's implications for small populations are direct. A rare disease cohort of several hundred patients, stratified by condition subtype, genotype, age at onset, geography of treatment, and institution of care, is re-identifiable by definition. Safe Harbor was designed for large populations in which stripping eighteen identifiers leaves enough ambiguity to protect the individuals within; rare disease cohorts of fewer than two hundred patients who share a known biological cause fall outside its operating envelope.

The Department of Justice has reached a parallel conclusion for a different reason. The [Final Rule on Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons](#), 28 CFR Part 202, implementing Executive Order 14117 and effective April 8, 2025, treats bulk access to American genomic data as a national security matter. The rule reflects a federal determination that genomic data cannot be meaningfully de-identified at scale and that its aggregation in adversarial hands is a strategic risk serious enough to override certain existing exemptions for clinical trial data sharing. The regulatory posture under which rare disease data is currently de-identified and sold predates both of these realities.

2.3 Consent Failures: The Canonical Cases

Two cases define the current consent landscape and together establish the limits of the de-identification-plus-consent model that the deidentified health data economy has relied on for two decades.

The first is Henrietta Lacks. Cells taken without her knowledge during cervical cancer treatment at Johns Hopkins Hospital in 1951 became the HeLa line, the most widely used cell line in biomedical research. Her descendants sued Thermo Fisher Scientific in 2021 over continuing commercial use of the line. On August 1, 2023, her 103rd birthday, the [parties announced a settlement](#) whose terms are confidential. The case did not resolve the underlying legal question of whether descendants have a property interest in tissues taken without consent from a deceased ancestor. It did establish that the question is now justiciable, seventy-two years after the initial extraction.

The second is 23andMe. The company filed for Chapter 11 protection in the Eastern District of Missouri in [March 2025](#). Regeneron entered an asset purchase agreement for approximately \$256 million in May, which was superseded at a reopened auction in June by a \$305 million bid from TTAM Research Institute, a nonprofit formed by co-founder Anne Wojcicki. During the pendency of the sale, twenty-seven states and the District of Columbia [filed suit in bankruptcy court](#) seeking to block the transfer of customer genetic data on the ground that consent given for one purpose, commercial ancestry analysis, could not support a different purpose, transfer of the data as a bankruptcy asset to whichever party prevailed at auction. The sale closed to TTAM in July 2025. The Consumer Privacy Ombudsman mechanism under which the sale was approved is now the working precedent for what happens to consumer genomic data when the original collecting entity exits the business.

The unifying lesson of the two cases is that the consent contract drawn at the time of collection does not survive the downstream events that determine how the data is actually used. Time, corporate reorganization, and scientific repurposing each exceed what a single-study consent document can accommodate. An architecture that produces reusable rare disease data at scale, under the Plausible Mechanism Framework or any successor framework, cannot rely on consent structures that fail at these predictable inflection points.

2.4 The Artificial Intelligence Substrate Problem

On January 7, 2025, the FDA issued draft guidance titled [Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations](#). The guidance introduces a Total Product Life Cycle framework for AI-enabled medical devices and requires manufacturers to assess training data for bias, document data provenance and inclusion and exclusion criteria, evaluate device performance across demographic subgroups, and maintain traceability from every data source used in training back to its origin. The comment period closed April 7, 2025.

The guidance addresses AI-enabled medical devices generally, and it is binding on rare disease practice by implication. Rare disease populations are the subgroups most likely to be absent from training data in the current data economy, and the traceability requirement cannot be satisfied at all by data whose collection custody has transferred through a bankruptcy auction or through a commercial aggregator who does not retain original consent records. The FDA is telling the field that data provenance is now a regulatory object, and the field is not in a position to comply with that requirement using the substrate it has.

The broader concern extends beyond regulated medical devices to the general use of clinical data as training corpus for medical artificial intelligence. The [OpenAI collaboration with Color Health announced in May 2024](#), expanded enterprise agreements between Microsoft and Epic, and similar arrangements throughout the sector reflect a rising appetite for clinical data as an input to model training. Aggregators that source data under HIPAA Safe Harbor de-identification can supply volume. They cannot supply the provenance, the active consent to training use, or the guaranteed representativeness of rare disease subgroups that the regulatory framework will increasingly require. A well-governed, fiduciary-held, longitudinal rare disease dataset is a materially different product from what the existing market supplies.

Part III. Cureledger: Fiduciary-Grade Privacy by Design

Cureledger is a platform for the custody, reuse, and monetization of human health data organized around a single architectural commitment. Patients retain a durable legal interest in the data they contribute; a fiduciary holds and administers that data under terms the patient sets; and every downstream transaction is provable against those terms.

The platform is designed first for rare disease, because rare disease is where the existing custody, consent, and data quality failures are most acute and where the commercial and scientific upside of resolving them is largest. The architecture generalizes to broader indications without modification.

3.1 The Life Data Trust

The core legal primitive of the Cureledger platform is the Life Data Trust. A contributor deposits data into a trust that exists as a distinct legal entity governed by trust law rather than by contract law. A trustee holds the data and administers it in accordance with terms the contributor specifies at the time of deposit and may revise subject to the trust's governance rules. The relationship is fiduciary. The trustee is bound by a duty of loyalty and a duty of care that run to the contributor as beneficiary, and those duties are enforceable through ordinary trust-law remedies.

The Life Data Trust replaces the click-through consent license as the instrument through which a patient participates in the data economy. A click-through license is a revocable, purpose-limited contract that binds the issuer only to the extent a court is willing to enforce it, and enforcement is difficult because the patient has typically received no consideration other than a service. A trust instrument creates an ongoing fiduciary obligation that survives the contributor's loss of capacity, the transfer of trustee, and the winding up of any operating entity. It is the legal form best suited to an asset, health data, whose value accrues over time and whose custody must persist across the contributor's lifetime and potentially beyond.

3.2 Default De-identification with Conditional Re-identification

Data in the Life Data Trust is de-identified by default using methods calibrated to the population size at issue. For small populations, Safe Harbor is insufficient and is supplemented or replaced by Expert Determination-grade statistical protection, cryptographic commitments that bind a record to a contributor without revealing identity, and zero-knowledge proofs that permit compliance validation without exposure of the underlying data. The result is a dataset in which any given record is provably compliant with the applicable regulatory regime, and in which the link from record to contributor is held only by the trustee under the terms of the trust.

A conditional re-identification path exists for one purpose, and that purpose is to inform the contributor. When a development in the research or clinical domain is relevant to an individual contributor, a matched therapy under the Plausible Mechanism Framework, for example, or a registry call from a sponsor initiating an external-control arm, the trustee may traverse the link from dataset record to contributor identity for the limited purpose of notifying that contributor. The contributor controls whether the link may be traversed at all, under what conditions, and for which categories of development. The default is that traversal is permitted only for developments the contributor has pre-authorized, and any traversal is logged and cryptographically attested.

This design resolves a tension that existing de-identification regimes leave unresolved. Full de-identification sacrifices the ability to inform the patient when their own contribution has produced a result that matters to them. Identified data sacrifices the privacy protection that makes large-scale reuse ethically and legally viable. The conditional re-identification path preserves both properties by placing the decision to traverse the link with the contributor and the execution of the traversal with the trustee.

3.3 Pro Rata Profit Participation

Revenue generated by authorized reuse of trust-held data flows back to contributors in proportion to their contribution. The distribution is administered by the trustee under the terms of the trust instrument and is verifiable against the cryptographic record of each reuse transaction. Contributors share directly in the revenue their data generates, as compensated beneficiaries of the trust under which it is held.

The pro rata mechanism serves three distinct purposes. It provides a direct financial incentive for contribution and for the longitudinal updates that make rare disease data exponentially more valuable over time. It establishes an ethical baseline that distinguishes Cureledger from aggregators operating under the Safe Harbor model. And it solves, in advance, the political problem that the Lacks and 23andMe cases expose: contributors whose data generates value have a structural claim on that value that the platform acknowledges from the outset rather than contesting after the fact.

3.4 Smart Contracts and Cryptographic Proofs as Verification Instruments

Cureledger uses smart contracts and cryptographic proofs to make the operation of the trust verifiable to parties who have reason to require verification: the contributor, the trustee, the regulator, and the purchaser of trust-held data. Smart contracts execute the terms of the trust where those terms can be reduced to code, notably the administration of pro rata distributions, the logging of reuse transactions, and the enforcement of contributor-defined access constraints. Cryptographic proofs, in particular zero-knowledge proofs, permit a party to establish a compliance-relevant fact without disclosing the underlying data or the underlying identity.

The product is the fiduciary relationship between contributor and trustee, together with the governance architecture that implements it. Smart contracts and cryptographic proofs provide the verification infrastructure for that relationship, rendering its operations auditable to contributors, regulators, and data purchasers without requiring those parties to trust one another or to trust Cureledger. Because the platform is legally a trust and technically a set of attested transactions, misuse is structurally harder than under the license-and-database architectures that dominate the market today.

3.5 Longitudinal and Sponsor-Independent by Construction

Because the data is held by a trust whose existence is independent of any operating entity, the longitudinal record survives the exit of any particular sponsor, researcher, or commercial partner. A clinical trial that ends when its sponsor discontinues the program leaves behind a dataset that other programs can access under the contributor's standing terms. A natural history study conducted under one foundation's auspices is available as external control data for subsequent programs without renegotiation of custody. Sponsors under this architecture are consumers of trust-held data, and custody of the data remains with the trust across the lifetime of any given program. The sponsor custody failure described in Part II is structurally prevented.

This property is what allows Cureledger to serve the Plausible Mechanism Framework as general-purpose infrastructure across programs and across years. The natural history data that the PMF relies on accumulates across programs and across years in a form that remains usable, verifiable, and legally reusable for the entire duration of the contributor's participation in the trust.

Part IV. Market and Business

4.1 The Deidentified Health Data Economy Today

The commercial deidentified health data market is large, concentrated, and growing. IQVIA, the incumbent, reported [fiscal year 2024 revenue of approximately \\$15.4 billion](#), a material share of which derives from its technology and analytics solutions segment built on non-identified patient records. [Datavant](#) operates the dominant patient-record tokenization and linkage layer following its 2021 merger with Ciox Health. [Truveta](#) aggregates data from a consortium of more than twenty health systems and, in 2024, launched the

Truveta Language Model trained on the consortium's records. [Komodo Health](#), [HealthVerity](#), [TriNetX](#), and [Flatiron Health](#) each serve adjacent segments of the same market. [Tempus AI](#) completed its initial public offering in June 2024 on NASDAQ under the ticker TEM, bringing a combined clinical, genomic, and imaging dataset to public market scrutiny.

Grand View Research and similar industry sources size the narrower real-world evidence solutions market at roughly \$2 billion in 2023 with projected growth to the range of \$6 to \$7 billion by 2030. The broader health data value stack that includes analytics, model training, and licensed derivative products is materially larger. Across the entire stack, patients participate as uncompensated inputs. The HIPAA Safe Harbor regime permits commercial aggregation and resale of de-identified data without further consent from or consideration to the individuals whose health histories constitute the underlying product.

4.2 Cureledger's Wedge

Rare disease is underserved by every major aggregator in the existing market. The small-N economics that make rare disease hard for pharmaceutical sponsors also make it hard for data aggregators whose business model relies on statistical depth within a single indication. A well-constructed rare disease dataset is a product the existing aggregators cannot easily produce, because producing it requires patient-level relationships, condition-specific consent, longitudinal follow-up of individuals rather than of events, and data standards compatible with natural history research. These are the properties Cureledger is built to deliver.

The Cureledger wedge is rare disease specifically and the initial commercial motion is focused accordingly. The Plausible Mechanism Framework supplies the regulatory demand: sponsors advancing individualized therapies under the PMF require natural history and external control data of exactly the sort trust-held rare disease data provides. Patient advocacy organizations and disease-specific foundations supply the contributor relationships: these are communities whose members are already organized, already motivated, and already contributing to fragmented registries that a fiduciary-grade trust can consolidate and properly steward.

4.3 Compounding Longitudinal Value

A rare disease dataset gains value with time along multiple independent axes. Volume grows as additional contributors join. Depth grows as existing contributors provide longitudinal updates covering disease trajectory, treatment response, and co-morbidity. Diversity grows as contribution expands across genotypes, subtypes, and geographies. Each of these axes multiplies with the others. A dataset that doubles in contributor count and doubles in per-contributor longitudinal depth does not merely quadruple in research value; it unlocks analyses that the earlier dataset could not support at any level of statistical power.

The compounding is especially pronounced for external control applications under the Plausible Mechanism Framework. A single well-characterized natural history cohort can serve as the reference for multiple successive therapy programs targeting the same condition, and each such reuse amortizes the cost of the original contribution across a widening base of beneficiaries. The appropriate reference point for trust-held rare disease data pricing is the downstream value a sponsor assigns to avoiding the construction of a bespoke comparator arm, which runs orders of magnitude above per-record rates in the commodity market.

4.4 Go to Market

The initial Cureledger market motion has three prongs, executed in parallel.

The first is partnership with rare disease foundations and patient advocacy organizations. These entities already hold contributor relationships and in many cases already administer registries. The Cureledger offer to them is a fiduciary-grade upgrade to their data custody posture, with pro rata return to their constituents, in exchange for migration of the existing registry into the Life Data Trust architecture.

The second is supply of external control and natural history data to pharmaceutical sponsors operating under the Plausible Mechanism Framework. The offer is a regulatory-grade dataset that satisfies the evidentiary requirements the draft guidance contemplates and that the sponsor cannot efficiently assemble on its own. Pricing reflects the saved cost of bespoke comparator construction rather than a per-record commodity rate.

The third is licensing to artificial intelligence laboratories requiring provenance-verified clinical data for model training. The offer is a training corpus whose inclusion criteria, contributor consent, and representativeness across rare disease subgroups are individually attested and cryptographically verifiable. This motion follows the regulatory adoption of the January 2025 FDA draft guidance on

artificial intelligence-enabled device software and the expected consolidation of provenance as a standard procurement requirement.

4.5 Generalization Beyond Rare Disease

The infrastructure that serves rare disease generalizes to any indication in which longitudinal depth, contributor consent, and provenance are meaningful. Chronic disease populations, oncology natural history cohorts, and cardiometabolic longitudinal studies all share the features that make rare disease the hardest case and that Cureledger's architecture accommodates by design. Generalization requires additional contributor onboarding and partner development; the engineering is already in place. The rare disease proof establishes the product; the larger-indication market establishes the commercial ceiling.

Part V. Moat

5.1 Patent Portfolio

Cureledger holds a portfolio of granted and filed patents covering dynamic data compliance controls, net-sum zero-knowledge proof formulations, cryptographic ciphers for personally identifiable and protected health information retrieval, compliance cipher deployment, blockchain-based health data management, metadata pointer management across data silos, data use permissions and royalty management, regulatory compliance automation for machine learning, and related subject matter. A fuller list is provided in Appendix B.

The relevance of the portfolio is defensive and enabling in roughly equal measure. The granted and pending claims cover the specific mechanisms by which compliance can be established, proved, and automated across regulatory regimes, and they accordingly raise the cost of entry for competitors attempting to replicate the fiduciary-grade data handling that Cureledger's trust architecture requires.

5.2 The Trust Architecture Itself

The Life Data Trust is a durable competitive asset independently of the patent portfolio. A trust is a legal form that imposes ongoing fiduciary duty on whoever administers it, and that duty is not replicable by a competitor who wishes to operate under a license-based consent model with weaker obligations. Competitors can declare fiduciary intent; they cannot incur fiduciary duty without accepting the architecture that creates it. The set of entities willing to accept that duty is small, the set of entities able to administer it competently is smaller, and the set that has already built the legal and technical infrastructure to do so is Cureledger.

5.3 Network Effects

Each program that reuses trust-held data strengthens the evidence base for the next. A natural history cohort used as external control in the first Plausible Mechanism Framework program for a given condition is more valuable when the second program begins, because it has been validated against a treatment arm and has accumulated an additional cycle of longitudinal contribution. The value to a later program is a function of how many earlier programs have already used and validated the same dataset. This is the classical network effect applied to longitudinal data, and it is the specific reason that an early lead in rare disease data custody compounds into a durable advantage across drug development generally.

Closing

The Plausible Mechanism Framework and the operational arrival of agentic artificial intelligence have, in the same twelve months, resolved two of the three constraints that held rare disease back. The third constraint, the absence of a fiduciary-grade, sovereign, longitudinal data layer, is an infrastructure and governance problem. Cureledger exists to do that work, for rare disease first and for human health data generally thereafter.

The people who live with rare diseases, and the families who care for them, have waited long enough for the system to produce cures. The audacity of the rare disease community has always been its insistence on cures in the face of economics that offered only lifelong therapy. The remainder of the field is now catching up. Cureledger is the infrastructure that closes the gap.

Appendices

Appendix A. Longitudinal Data Value Model

To be populated. Plain-language version of the v1 growth model. Parameters: contributor count, per-contributor longitudinal depth, subtype and genotype diversity, reuse frequency. Output: dataset value as a function of time. Token-reward and validator-reward mathematics removed.

Appendix B. Patent Portfolio

To be populated with verified current counts. Categories from v1: data compliance across regulatory regimes, data exchange compliance, data authentication signatures, data use permissions and royalty management, metadata pointer management, healthcare use case inventions, contracting and finance inventions, regulatory compliance machine learning, blockchain consensus, decentralized web3 compliance automation.

Appendix C. References

Inline hyperlinks throughout the body resolve to primary sources. Full bibliography to be compiled at finalization. Primary citation set:

- FDA, *Considerations for the Use of the Plausible Mechanism Framework*, Docket FDA-2026-D-1256, Feb. 23, 2026.
- FDA, *Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations*, Jan. 7, 2025.
- FDA, *Rare Diseases: Considerations for the Development of Drugs and Biological Products*, final guidance, Dec. 2023.
- FDA, *Rare Diseases: Natural History Studies for Drug Development*, draft guidance, March 2019.
- Kim et al., *Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease*, N. Engl. J. Med., Oct. 2019 (Milasen).
- Rocher, Hendrickx, and de Montjoye, *Estimating the success of re-identifications in incomplete datasets using generative models*, Nat. Commun. 10:3069 (2019).
- U.S. Dep't of Justice, *Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons*, 28 C.F.R. Part 202, effective Apr. 8, 2025.
- *Lacks v. Thermo Fisher Scientific*, D. Md., settled Aug. 1, 2023.
- *In re 23andMe Holding Co.*, Bankr. E.D. Mo., filed March 2025; sale to TTAM Research Institute approved July 2025.
- IQVIA Holdings Inc., Form 10-K for fiscal year 2024.
- Tempus AI, Inc., Form S-1 and subsequent filings, June 2024 IPO.
- Grand View Research, *Real World Evidence Solutions Market*, 2024 edition.