Follow the Data: Provenance Infrastructure in Regulated-AI Acquisitions

A diligence, valuation and integration framework for data lineage, evidence continuity and switching-cost claims

Authored by Chennakeshav Adya

Independent Researcher

September 2026

Abstract

Regulated artificial-intelligence systems depend on more than a model. They depend on evidence connecting source data, transformations, training runs, evaluations, approvals, deployments and consequential decisions. In an acquisition, that chain can be a valuable operating asset or an expensive illusion. A buyer needs to know whether provenance is complete enough to support compliance and change control, portable enough to survive integration, and embedded deeply enough to create genuine customer switching costs.

This paper develops the Provenance Infrastructure Acquisition Framework. It applies the NIST AI Risk Management Framework, the W3C PROV family, the European Union AI Act, C2PA technical specifications, software supply-chain standards and current data-governance practice to transaction diligence. The framework separates six layers: source identity; lawful and contractual rights; transformation history; model and software lineage; evaluation and approval evidence; and production decision records. Each layer has distinct control, integration and valuation implications.

The central conclusion is that provenance creates defensible value when it is reliable, queryable, exportable and connected to accountable operating decisions. Deep integration can support retention and expansion because replacing the platform would require customers to reconstruct trusted histories, control mappings and evidence workflows. Proprietary formats, incomplete lineage and unverifiable records create lock-in without durable value. The buyer should therefore value collected revenue and demonstrated customer dependence while funding remediation, interoperability and evidence preservation.

Four tables and three figures convert these principles into an acquisition method. A hypothetical valuation bridge illustrates how evidence quality, integration depth, concentration and remediation affect value. All numerical values are illustrative management assumptions prepared solely to demonstrate the framework; they are not market observations, forecasts or valuation advice.

JEL Classification: G31, G32, G34, K22, L86, O32

Keywords: data provenance, AI lineage, regulated AI, M&A due diligence, switching costs, data governance, audit trail, integration

This paper is an educational and structural analysis prepared for research purposes. It is not investment advice, an offer, or a solicitation. It contains no client information.

1. DEFINE THE INVESTMENT DECISION

The investment committee must decide whether the target owns a trusted evidence infrastructure or a collection of metadata features. The distinction affects revenue quality, integration cost, regulatory exposure and the credibility of switching-cost claims. Diligence should begin with the decisions that customers use the platform to support, the evidence required for those decisions and the consequences of an incomplete record.

The transaction thesis should state which provenance functions drive retention, which are replaceable utilities and which depend on services or manual work. It should identify the legal entities, systems and people responsible for the record. A broad claim that the platform provides responsible AI cannot support valuation without this operating map.

2. DEFINE PROVENANCE

The W3C describes provenance as information about entities, activities and people involved in producing data or another thing. Its PROV family provides concepts and interoperable representations for recording generation, use, derivation, attribution and association. These concepts are useful because AI evidence crosses organisational and technical boundaries.

For acquisition purposes, provenance is the supported history of an artefact and the actions that changed, evaluated, approved or used it. The record should answer what existed, where it came from, who acted, what process occurred, when it occurred and which later outputs depended on it. A timestamp alone is not provenance.

3. IDENTIFY THE REGULATED USE

Regulatory relevance depends on the system, jurisdiction and intended use. The EU AI Act requires data governance, technical documentation and record-keeping for high-risk AI systems. Other regulated sectors impose model-risk, privacy, safety, cybersecurity, records and audit obligations. The buyer should map the target's customer use cases to current requirements with qualified legal and regulatory advisers.

The map should distinguish a platform that stores evidence from one that performs a regulated assessment or makes a compliance claim. Software can support an obligation without assuming the customer's legal responsibility. Contracts, product language and actual practice should remain aligned.

4. MAP THE EVIDENCE CHAIN

The evidence chain begins before model training. It includes source systems, collection authority, consent or licence, extraction, labelling, filtering, transformation, feature engineering, training configuration, code, model weights, evaluations, approvals, deployment, monitoring and downstream decisions. External models and data introduce additional chains controlled by third parties.

Diligence should select material outputs and trace them backwards to source. It should also select source records and trace their downstream use. Bidirectional testing reveals orphan records, silent transformations and undocumented dependencies. The result becomes the base map for technical, legal and commercial workstreams.

5. SEPARATE SIX PROVENANCE LAYERS

The first layer identifies the source. The second records legal and contractual rights. The third records transformation history. The fourth links data to model and software versions. The fifth records evaluation, review and approval. The sixth connects deployed outputs to operating decisions and later outcomes.

Completeness at one layer cannot cure absence at another. A buyer may know which dataset trained a model while lacking evidence of permitted use. It may know the licence while lacking a reproducible transformation. The diligence conclusion should score each layer separately and then examine the chain as a whole.

Table 1. Provenance layers and acquisition evidence

Layer	Core question	Direct evidence	Principal acquisition risk
Source identity	What data, model, code or content entered the chain?	Source identifiers, manifests, hashes, supplier records and acquisition logs	Unknown origin prevents reproducibility and reliable rights analysis
Rights and restrictions	Could the artefact be collected, used, shared and transferred?	Contracts, licences, consent records, notices, policy decisions and restrictions	Change of control or new use may be prohibited
Transformation history	What changed and why?	Versioned pipelines, parameters, code commits, lineage events and quality results	Undocumented processing makes outputs irreproducible
Model and software lineage	Which components produced the output?	Model registry, software bill of materials, dependencies, weights and deployment records	Vulnerable or unlicensed components enter the transaction
Evaluation and approval	Who tested, reviewed and authorised release?	Test data, metrics, thresholds, exceptions, sign-offs and change records	Performance claims lack an accountable basis
Production decisions	How was the output used and what followed?	Inference logs, workflow events, human decisions, overrides and outcomes	The buyer cannot investigate harm or monitor effectiveness

The evidence required varies by system, jurisdiction and intended use; the table is a transaction framework rather than legal advice.

6. TEST SOURCE IDENTITY

Every material artefact should have a stable identifier and sufficient context to distinguish it from later versions. Hashes can support integrity, while manifests and registries provide meaning. Source identity should survive transfer between storage, development and production environments.

The buyer should test whether the same identifier resolves consistently across systems. It should inspect duplicate, missing and recycled identifiers. If customers create their own identifiers, the platform should preserve tenant boundaries and mappings. A lineage graph that merges distinct sources can produce convincing but incorrect histories.

7. TEST RIGHTS AND RESTRICTIONS

Rights provenance connects an artefact to the legal basis for collection, processing, training, evaluation, distribution and transfer. Relevant evidence can include contracts, licences, consent, notices, public-sector permissions and internal approvals. Restrictions may depend on geography, purpose, retention, onward transfer or model type.

The buyer should trace a sample of material datasets and models to executed documents. It should identify change-of-control provisions and obligations that follow the data. Management assertions should remain separate from verified rights. A technical record without rights evidence may increase the speed at which a problem can be located without making the use lawful.

8. TEST TRANSFORMATION HISTORY

Transformation provenance should show how raw inputs became the data used by a model or decision process. It should record code, parameters, environment, operators, timing, quality tests and exceptions. Reproducibility requires more than a current pipeline because historical outputs may have been created under earlier versions.

Diligence should rerun selected transformations where feasible and compare outputs. It should inspect manual spreadsheets, notebooks and emergency changes outside controlled pipelines. Material differences should be explained and recorded. The cost of reconstructing undocumented history belongs in the transaction model.

9. TEST MODEL LINEAGE

Model lineage links training data, code, configuration, base model, fine-tuning, weights, evaluations and deployments. A model name is insufficient when multiple versions or supplier updates exist. The record should also identify prompts, retrieval sources, guardrails and orchestration components where they materially affect behaviour.

The buyer should choose production outputs and identify the exact model stack that generated them. It should confirm that historical versions remain available or that their essential evidence is preserved. A platform that overwrites versions cannot support reliable incident analysis.

10. TEST SOFTWARE LINEAGE

AI systems depend on libraries, containers, APIs, drivers and infrastructure. Software bills of materials can identify components, yet they require accurate build and deployment links. SPDX and CycloneDX provide structured formats that can support exchange across organisations.

Diligence should compare declared components with build files, container images and production scans. It should inspect known vulnerabilities, licences, end-of-life dependencies and supplier concentration. Software lineage should connect a component finding to the affected models, deployments and customers.

11. TEST EVALUATION LINEAGE

Evaluation lineage records the dataset, method, metric, threshold, environment, model version, reviewer and result. It should distinguish development testing from independent validation and production monitoring. A summary score without test-population provenance provides limited evidence.

The buyer should examine whether benchmark data entered training, whether labels changed and whether subgroup results were preserved. It should identify exceptions approved outside policy. Evaluation evidence should allow another competent person to understand what was tested and what the result can support.

12. TEST APPROVAL LINEAGE

Approval lineage shows who authorised a dataset, model, threshold, deployment or change and under which authority. The record should include decision criteria, conflicts, exceptions and conditions. Workflow completion by an unauthorised user is not approval.

Diligence should trace samples through role definitions, identity systems and signed records. It should inspect shared accounts, delegated authority and retrospective approvals. When approval migrated from email to platform workflows, the historical archive should remain connected.

13. TEST PRODUCTION LINEAGE

Production lineage connects an output to the deployed system, input context, time, location and downstream action. Logging should be proportionate to the use and compatible with privacy, security and retention requirements. The EU AI Act includes record-keeping requirements for high-risk systems and requires automatic logging capability over the system lifetime.

The buyer should confirm that logs are generated, retained, protected and queryable. It should examine gaps during outages, migrations and supplier incidents. A product feature that can create logs provides limited value when customers do not enable or retain them.

14. CONNECT PROVENANCE TO ACCOUNTABILITY

Provenance supports accountability by making actions and dependencies examinable. It does not itself allocate responsibility. Contracts, governance, authority and law determine who must act on the evidence. The platform should therefore connect records to owners and escalation paths.

The buyer should identify who investigates lineage exceptions, who approves remediation and who communicates with customers or authorities. It should test actual incidents rather than policy alone. Unresolved exceptions reveal whether the system is an operating control or a passive archive.

15. APPLY THE NIST AI RMF

The NIST AI Risk Management Framework organises activity through Govern, Map, Measure and Manage. NIST states that maintaining training-data provenance and supporting attribution of decisions to subsets of training data can assist transparency and accountability. Provenance should therefore support risk decisions across the lifecycle.

Diligence should map product functions and customer workflows to the four functions. It should identify evidence produced at each step and decisions that remain outside the platform. Coverage should be assessed against the customer's material risks rather than the quantity of stored metadata.

16. APPLY W3C PROV

W3C PROV centres on entities, activities and agents, together with relationships such as generation, use, derivation, attribution and association. This provides a useful interoperability test. A platform should represent its internal concepts in a way that can be exported and understood beyond one interface.

The buyer should request a complete export for selected records and map it to a standard model. It should assess identifiers, semantics, bundles and links across systems. Proprietary enhancements can add value while preserving a portable core.

17. APPLY THE EU AI ACT

For high-risk systems, the EU AI Act addresses data and data governance, technical documentation and automatic record keeping. Annex IV specifies information concerning versions, interactions, capabilities, limitations, input data, monitoring and control. These requirements create a practical demand for connected evidence.

The buyer should test whether the platform can support the customer's required documentation and logs using verified records. Product claims should distinguish support from compliance. Regulatory classification and obligations require case-specific advice, and the acquirer should avoid recognising unsupported compliance revenue.

18. EXAMINE C2PA CONTENT CREDENTIALS

C2PA provides technical specifications for content provenance and authenticity through signed manifests and assertions. It can help record origin and edits for digital content. Its scope differs from full AI-system lineage, which includes data rights, transformations, models, evaluations and decisions.

Diligence should identify where C2PA or comparable credentials enter the chain and how verification results are handled. A valid credential can support integrity and attribution for represented claims. It cannot establish that every claim is true or that the content is suitable for a regulated decision.

19. DISTINGUISH INTEGRITY FROM TRUTH

Cryptographic signing can show that a record has not changed after signing and identify the signing key. It does not prove that the original assertion was accurate, complete or authorised. The transaction thesis should separate record integrity from evidence reliability.

The buyer should inspect identity proofing, key management, revocation, timestamping and verification. It should also examine how false or mistaken assertions are corrected without erasing history. A strong signature over weak evidence remains weak evidence.

20. ESTABLISH THE DILIGENCE POPULATION

The diligence population should include material customers, regulated use cases, datasets, models, environments, suppliers, evidence types and incidents. Sampling should reflect revenue, risk, novelty and integration dependence. Management-selected demonstrations should not define the population.

The buyer should reconcile platform records to contracts, invoices, product telemetry, support tickets and customer confirmations where permitted. Differences reveal unused licences, services dependence and incomplete coverage. The reconciled population becomes the denominator for commercial and technical conclusions.

21. BUILD THE EVIDENCE MATRIX

The evidence matrix should connect every transaction question to a direct source, owner, period, test and conclusion. It should distinguish verified records, management estimates and unresolved items. This prevents a polished demonstration from substituting for population evidence.

High-priority questions include source coverage, rights, transformation reproducibility, model versioning, customer use, exportability, incidents and remediation. Each conclusion should identify the record that supports it and the limitation that remains. The matrix should continue into integration planning.

Table 2. Provenance diligence evidence matrix

Diligence question	Primary evidence	Test	Decision consequence
Does the platform capture material lineage?	Reconciled artefact population and event records	Trace outputs backwards and sources forwards	Scope remediation and challenge product claims
Are records reliable?	Signed events, identities, timestamps, audit logs and change history	Reperform selected queries and inspect exceptions	Adjust reliance and representations
Are rights transferable?	Executed licences, consents, supplier terms and restrictions	Trace sampled artefacts to documents and change-of-control terms	Exclude assets or condition close
Is customer dependence demonstrated?	Usage, workflow integrations, renewals, support records and interviews	Compare integration depth with retention and expansion	Support or reduce switching-cost value
Can the chain move to the buyer?	Export schemas, APIs, documentation and migration tests	Export and rebuild selected histories	Fund migration or preserve a separate stack
Can incidents be investigated?	Case files, affected-record queries, corrective actions and notices	Reconstruct selected incidents end to end	Adjust liability, reserves and integration controls

Sampling should be designed for the target's risk and customer mix; percentages and thresholds below are not prescribed.

22. MEASURE COMPLETENESS

Completeness should be measured against required events and relationships, not total record volume. A platform can store billions of events while missing the approval that gives a deployment authority. The buyer should define a required chain for each use case and calculate supported coverage.

Measures can include artefacts with verified source identity, transformations linked to code, deployments linked to approved models and decisions linked to outputs. Missingness should be segmented by customer, product version and time. Improvement in recent cohorts does not repair historical exposure.

23. MEASURE ACCURACY

Accuracy asks whether the recorded relationship reflects what occurred. Automated instrumentation may be more consistent than manual entry, yet configuration errors can scale across the estate. Independent reconciliation is required.

The buyer should compare lineage records with build logs, storage events, cloud configurations, model registries and operating decisions. It should investigate impossible sequences and duplicate identifiers. Accuracy conclusions should state the tested period and population.

24. MEASURE TIMELINESS

Provenance recorded at the time of activity is generally more useful than retrospective reconstruction. Delay can weaken incident response and allow unauthorised changes to enter production. The platform should record event time, ingestion time and later correction.

Diligence should calculate latency for material events and identify offline systems. It should examine whether customers can alter timestamps or backfill records without preserving the original. Transaction value depends on the history that existed before diligence, not a reconstruction created for sale.

25. MEASURE QUERYABILITY

Evidence creates operating value when authorised users can retrieve it for a decision, audit or incident. Queryability depends on identifiers, indexes, semantics, access control and performance. Dashboards alone may not support complex dependency questions.

The buyer should run representative queries such as every deployment affected by a dataset, every customer using a vulnerable component and every decision based on a withdrawn model. It should record time, manual effort and false results. Query performance should be tested at actual scale.

26. TEST INTEROPERABILITY

Interoperability allows provenance to cross tools, organisations and transaction boundaries. Standards can help, but field mappings, identity resolution and semantics remain important. The platform should document exports and preserve stable identifiers.

Diligence should export a selected chain into an independent environment and verify that meaning survives. It should identify proprietary fields and transformations. A buyer can credit differentiated analytics without accepting an opaque record that cannot be moved or examined.

27. TEST PORTABILITY

Portability addresses whether a customer or acquirer can retrieve complete records in a usable format. It affects exit rights, regulatory response and business continuity. Portability can reduce coercive lock-in while strengthening trust and adoption.

The buyer should inspect contractual export rights, fees, timing, formats and support obligations. It should perform a migration test. Exporting flattened events without relationships, schemas or evidence attachments may satisfy a narrow data extract while failing to preserve the chain.

28. DISTINGUISH SWITCHING COST FROM CAPTIVITY

A defensible switching cost arises because the platform is embedded in valuable workflows and preserves trusted history that would be costly to recreate. Captivity arises when poor portability, withheld data or proprietary obscurity blocks exit. The first can support durable value; the second can create regulatory, litigation and reputation risk.

The buyer should identify the work required to replace the platform and the reasons for that work. It should separate configuration, integrations, validated methods and historical evidence from contractual penalties or unusable exports. Customer interviews and actual migrations provide stronger evidence than sales claims.

29. SEGMENT INTEGRATION DEPTH

Integration depth can be classified as record, workflow, control and decision integration. Record integration stores evidence. Workflow integration coordinates tasks. Control integration enforces gates. Decision integration supports approvals, incidents and external reporting. Each level has different implementation cost and retention implications.

The buyer should classify customers using verified telemetry and configurations. Contracted modules are not equivalent to operating use. Revenue cohorts should be analysed by depth, tenure, industry and service burden.

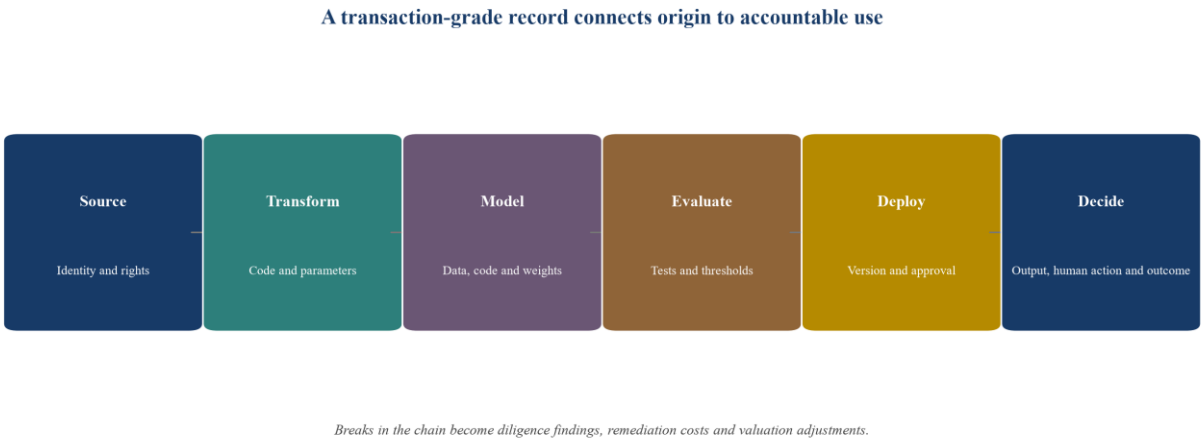

Figure 1. The end-to-end provenance chain

Every link requires identity, time, authority and preserved relationships; the accountable decision remains with the authorised organisation and people.

30. ANALYSE REVENUE COHORTS

Revenue cohorts should separate customers by integration depth and observed use. The buyer should compare gross retention, expansion, collected revenue, implementation time, support cost and contribution. It should examine whether deeper customers stay because the platform creates value or because migration is unusually difficult.

The analysis should control for tenure, contract length, sector and customer size. Small samples require caution. Management-estimated allocations should be labelled and reconciled where possible. Cohort evidence can support switching-cost value only when the definitions and records are reliable.

Table 3. Illustrative revenue cohorts by integration depth

Cohort	Illustrative annual collected revenue	Illustrative gross retention	Illustrative support cost as revenue	Diligence interpretation
Record integration	USD 2.0m	82%	14%	Useful archive, but replacement and export may be manageable
Workflow integration	USD 3.0m	90%	18%	Embedded tasks support retention, subject to implementation quality
Control integration	USD 3.5m	95%	21%	Gates and evidence mappings raise value and operational dependency
Decision integration	USD 1.5m	97%	27%	Strong dependence may coexist with concentrated expertise and service cost

Every value is a hypothetical management assumption for framework demonstration; no value is a market observation or forecast.

31. TEST RECURRING REVENUE

Recurring labels should be reconciled to executed contracts, invoices and collections. Platform subscriptions may include non-recurring implementation, evidence migration, control mapping and managed services. Renewals may depend on specialist teams rather than product use.

The buyer should calculate recurring collected revenue by customer and component. It should identify minimum commitments, usage charges, termination rights and renewal events. Services required to maintain the provenance chain belong in the cost-to-serve analysis.

32. ANALYSE CUSTOMER CONCENTRATION

Regulated-AI provenance can develop around a small group of early customers. Concentration increases commercial and product risk because one customer's requirements may shape the architecture. The buyer should examine revenue, evidence volume, product customisation and reference dependence.

Contracts should be reviewed for most-favoured terms, development obligations and ownership of custom controls. A large customer can validate the product while constraining future reuse. The valuation should reflect both verified demand and concentration exposure.

33. TEST CUSTOMER OUTCOMES

Customer value should be evidenced through reduced evidence-collection time, faster investigations, fewer control failures, improved audit readiness or supported approvals. The

buyer should distinguish measured outcomes from testimonials. Baselines and denominators matter.

Diligence should inspect customer records and, where permitted, conduct reference calls. It should ask what decision changed, which alternative existed and whether the platform remained essential after implementation. Outcomes support willingness to pay more directly than feature counts.

34. ANALYSE IMPLEMENTATION ECONOMICS

Implementation can involve connectors, data mapping, identity resolution, control configuration, historical migration and validation. The buyer should measure internal and partner hours, cloud cost, rework and elapsed time. Capitalised implementation should be reconciled with accounting policy.

Standard templates can improve contribution, while regulated use cases may remain bespoke. The model should not assume that historic low pricing will support future margins. Integration backlog and customer commitments should enter working-capital and capacity planning.

35. ANALYSE SUPPORT ECONOMICS

Support can include technical operations, evidence interpretation, regulatory mapping and incident response. These activities have different skills and margin profiles. Ticket counts without severity and effort provide limited insight.

The buyer should connect support demand to product version, customer depth and data volume. It should identify customers dependent on named specialists. Automation can reduce routine work when resolutions are verified and controlled. Unrecorded expert effort can make platform margins appear stronger than the operating reality.

36. EXAMINE SUPPLIER DEPENDENCE

The target may depend on cloud platforms, model providers, identity services, signing authorities, data catalogues and compliance content. Provenance should record these dependencies and relevant versions. Supplier changes can alter the evidence chain without a customer-visible product release.

Diligence should inspect contracts, service levels, change notices, concentration and exit plans. It should test whether third-party outages or discontinued APIs create evidence gaps. The transaction model should fund transition where a dependency conflicts with the buyer's architecture or risk appetite.

37. EXAMINE DATA RESIDENCY

Provenance records can contain personal, confidential, security-sensitive and legally privileged information. Storage and cross-border transfer must be designed for the applicable data and jurisdictions. Replicating complete histories can increase exposure.

The buyer should map storage, backups, support access and subprocessors. It should test tenant segregation and deletion. Integration planning should preserve required evidence while applying minimisation, retention and transfer controls. Legal analysis should be specific to the data and use.

38. EXAMINE RETENTION AND DELETION

Evidence obligations and privacy requirements can pull in different directions. The platform needs policies that identify records, purposes, periods, holds and authorised disposal. Deleting a source artefact while preserving an unexplained reference can make the chain unusable.

Diligence should test retention rules against actual storage and backups. It should inspect legal holds, customer-configured periods and deletion certificates. Transaction migration should avoid resetting retention clocks or silently extending storage.

39. EXAMINE SECURITY

Provenance infrastructure can become a high-value map of systems, data and weaknesses. Security should cover identity, least privilege, encryption, tenant isolation, logging, vulnerability management and incident response. Signing keys require special protection.

The buyer should examine independent tests, incidents and remediation. It should determine whether administrators can alter records without detection. Immutable design claims should be tested against privileged access, backups and correction workflows.

40. EXAMINE RESILIENCE

The provenance service should preserve records and continue critical capture during failures. Recovery objectives should reflect customer use and evidence obligations. Backups are insufficient if schemas, identifiers or signing keys cannot be restored consistently.

Diligence should inspect architecture, exercises and actual outages. It should test event buffering, replay and duplicate handling. The integration plan should protect chronology and relationships during migration, failover and provider change.

41. EXAMINE INCIDENT RESPONSE

Incident response should use provenance to identify affected artefacts, deployments, customers and decisions. The record should show investigation steps, evidence preserved, approvals, communications and corrective action. Speed without reliable scope can worsen an incident.

The buyer should reconstruct selected events from detection to closure. It should compare the case file with platform records and customer notices. Repeated gaps may indicate product limitations or weak operating discipline. Open incidents should enter the transaction risk register.

42. EXAMINE CORRECTION AND REVOCATION

Provenance records sometimes require correction when an assertion, identity or relationship is wrong. The system should preserve the original, reason, authorised correction and downstream impact. Revoked credentials or withdrawn models should remain discoverable.

Diligence should test whether corrected records propagate to reports and alerts. It should inspect key and credential revocation. An append-only claim should not prevent accountable correction; the design should preserve history while making current status clear.

43. EXAMINE AUDITABILITY

Auditability requires readable evidence, defined control ownership and reproducible queries. A complex graph can remain unauditable when meanings change without version control. The platform should preserve schemas, policy versions and control mappings.

The buyer should ask an independent team to reproduce selected reports from source records. Differences should be explained. Audit exports should include limitations and query parameters. Screenshots provide weak long-term evidence when the underlying data and logic are unavailable.

44. EXAMINE LEGAL HOLDS AND DISCOVERY

Litigation, investigations and regulatory requests may require preservation and production. The platform should identify custodians, artefacts, changes and chain of custody. Search and export should respect scope, privilege and access controls.

Diligence should inspect hold capabilities, actual requests and response records. It should identify data that cannot be collected or produced within required times. The transaction should preserve holds across systems and entities. Qualified counsel should direct case-specific decisions.

45. TEST MANAGEMENT CLAIMS

Management may claim complete lineage, immutable records, regulatory compliance, rapid deployment or high switching costs. Each claim should be translated into a testable definition, population and period. Unsupported adjectives should remain outside the investment case.

The buyer should retain the original claim, evidence and conclusion. Where evidence covers only a subset, the conclusion should state that subset. This discipline protects the investment committee from converting a demonstration into a portfolio-wide fact.

46. QUANTIFY REMEDIATION

Remediation can include source reconciliation, rights review, connector rebuilds, schema mapping, identity cleanup, historical backfill, security work and customer migration. Cost should include specialist time, customer disruption and parallel operation.

The buyer should estimate by work package and dependency. Historical reconstruction may remain impossible for some records, requiring disclosure, contractual treatment or exclusion. A single percentage reserve is insufficient when issues have different timelines and consequences.

47. BUILD THE VALUATION BRIDGE

The valuation bridge should begin with verified revenue and contribution. It can then recognise demonstrated retention, expansion and software leverage. Adjustments should reflect concentration, services dependence, incomplete lineage, rights uncertainty, security findings, portability and integration cost.

The bridge should not capitalise hypothetical regulatory demand without customer evidence. Synergies should have owners, cost and timing. The result should show which value depends on closing conditions and post-close execution.

Table 4. Hypothetical provenance-infrastructure valuation bridge

Item	Illustrative value effect	Evidence required	Treatment
Verified recurring collected revenue and contribution	USD 42.0m enterprise-value base	Contracts, invoices, collections, product use and cost allocation	Base case
Demonstrated control and decision integration	+USD 6.0m	Cohort retention, workflow evidence and customer references	Conditional uplift
Reusable mappings, connectors and evidence models	+USD 3.0m	Ownership, reuse, maintenance cost and paid deployment	Conditional uplift
Customer concentration and key-person dependence	-USD 4.0m	Revenue concentration, staffing and succession evidence	Risk adjustment
Rights and historical-lineage gaps	-USD 5.0m	Sample testing, legal review and remediation estimate	Risk adjustment
Migration, interoperability and security programme	-USD 3.5m	Costed integration plan and independent findings	Funded adjustment
Illustrative adjusted enterprise value	USD 38.5m	Completion of all evidence gates	Decision output

All values are illustrative management assumptions prepared only to demonstrate the framework; they are not market observations, a valuation opinion or a forecast.

48. AVOID DOUBLE COUNTING

The same customer dependence can appear in retention, pricing, growth and a strategic premium. The buyer should identify each causal driver and credit it once. Software leverage may already be reflected in contribution.

The bridge should separate target value from buyer-specific synergy. Historical evidence should remain distinct from forecasts. Sensitivities should show the effect of lower retention, higher remediation and delayed integration. This produces a more decision-useful range.

49. DESIGN REPRESENTATIONS AND WARRANTIES

Transaction protections can address ownership, licences, permitted use, security, records, customer claims, compliance and disclosed incidents. Drafting should follow verified facts and qualified legal advice. Broad assurances cannot replace diligence.

Schedules should identify material datasets, models, suppliers and restrictions. Knowledge qualifiers, baskets, caps and survival periods affect risk allocation. Specific indemnities or escrows may be appropriate for identified issues. Operational remediation should continue regardless of contractual recovery.

50. SET CLOSING CONDITIONS

Closing conditions can require critical rights consents, security remediation, evidence preservation, customer approvals, key-person arrangements and delivery of complete exports. Conditions should be objective and verifiable.

The buyer should avoid conditions that depend on newly created histories presented without independent support. A remediation can improve controls prospectively while leaving historical limitations. The completion memorandum should state both results and remaining limitations.

51. PLAN DAY ONE

Day One should preserve service, identities, keys, logs, holds, customer access and incident escalation. Product rebranding and data migration should follow evidence preservation. The combined group should know which entity remains responsible for existing commitments.

The plan should name owners and fallback actions. It should prohibit uncontrolled connector changes and record deletions. Customers with regulated workflows may need advance communication or approval. Business continuity should be demonstrated before cutover.

52. DESIGN THE TARGET ARCHITECTURE

The target architecture should define systems of record, identifiers, event models, interfaces, access, retention and reporting. It should decide which proprietary capabilities remain and which functions move to group platforms. Interoperability should be a design requirement.

The buyer should preserve source and transformation histories during consolidation. A data lake containing copied events is not equivalent to a provenance system unless relationships and meaning survive. Architecture decisions should be tested with representative end-to-end chains.

53. SEQUENCE MIGRATION

Migration should proceed by evidence domain or customer cohort with parallel verification. Each wave should define entry criteria, reconciliation, exception handling, customer acceptance and rollback. High-risk or legally constrained histories may remain on the original platform longer.

The buyer should measure record counts, relationship integrity, query equivalence and access. Manual exceptions require owners. Completion should depend on preserved meaning and operating use rather than copied bytes.

54. PROTECT CUSTOMER TRUST

Customers may have selected the target because it provides an independent evidence layer. Acquisition by a model, cloud or consulting provider can change perceived independence. The buyer should assess conflicts, access and product governance.

Communication should explain continuity, controls and data use accurately. Contract changes should be explicit. Trust can strengthen when the acquirer funds resilience and interoperability, provided customers retain appropriate control of their evidence.

55. GOVERN PRODUCT CHANGES

New features should identify affected evidence, schemas, control mappings and customer commitments. Change approval should include legal, security, technical and customer-impact review. Generated mappings or summaries require validation.

The combined company should maintain versioned documentation and migration paths. Deprecation should consider record retention and reproducibility. Product velocity becomes valuable when it preserves the chain and reduces reliable customer effort.

56. GOVERN AI USED INSIDE THE PLATFORM

The provenance platform may itself use AI to classify artefacts, map controls, detect anomalies or summarise evidence. Those functions need their own lineage, evaluation and human accountability. A system cannot support trusted evidence while obscuring how its own material outputs were produced.

Diligence should identify AI-assisted functions and consequences. It should inspect training or retrieval sources, tests, thresholds, overrides and monitoring. Generated content should be distinguishable from verified records.

57. MONITOR POST-CLOSE PERFORMANCE

The board dashboard should connect revenue and product use with evidence quality. Measures can include collected recurring revenue, customer depth, complete chains, unresolved exceptions, query time, export success, incidents, support effort and remediation progress.

Definitions should remain stable and sources documented. Aggregate completeness can hide gaps in a material customer or regulated use. The dashboard should distinguish verified records from management estimates and identify accountable owners.

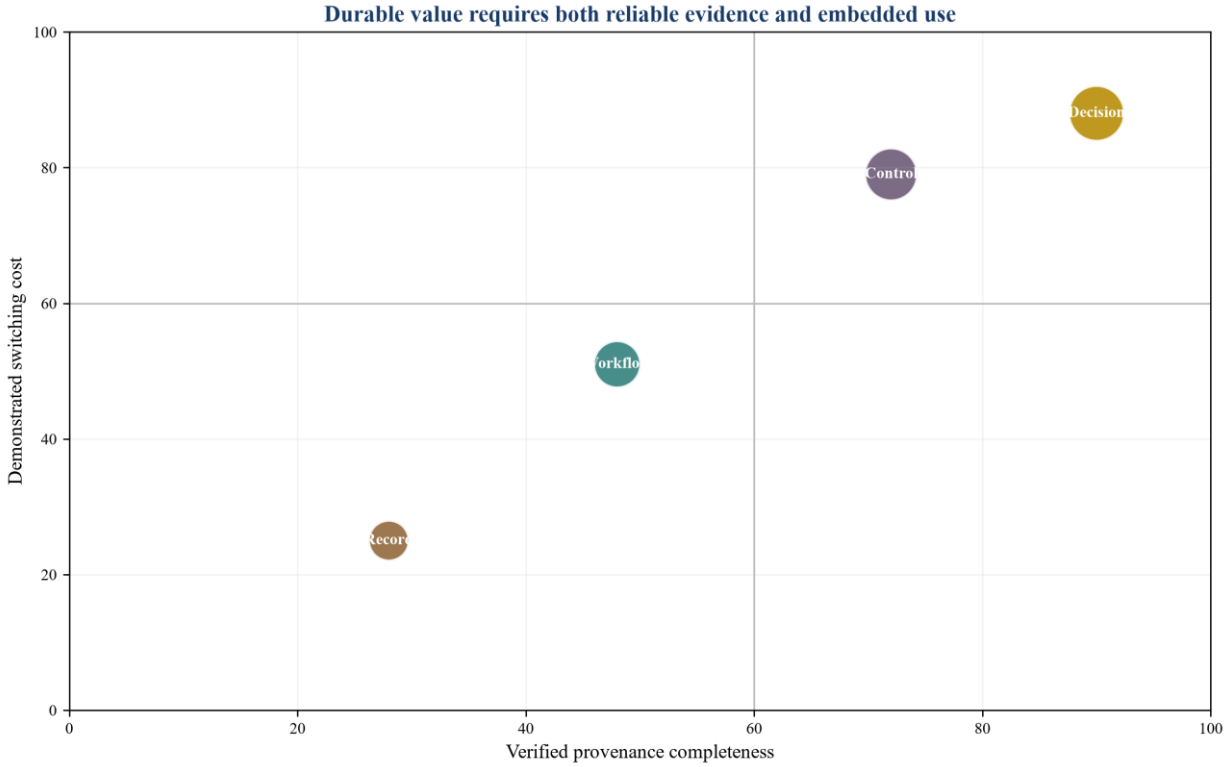

Figure 2. Provenance completeness and switching-cost profile

The positions and scores are illustrative; a buyer should populate the chart with verified customer evidence.

58. APPLY FIVE INVESTMENT GATES

The first gate tests rights and source identity. The second tests lineage completeness and accuracy. The third tests customer use and economics. The fourth tests portability, security and resilience. The fifth tests integration and accountability.

An open gate should trigger evidence work, remediation, a condition, repricing or a stop decision. The investment committee should record the evidence, owner and deadline. Closing a gate requires a verified result rather than an intention.

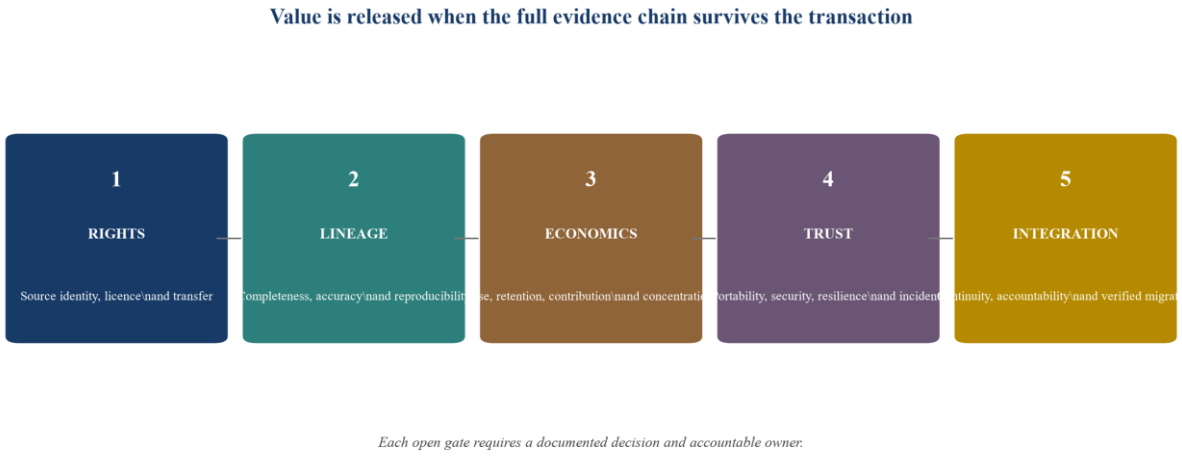

Figure 3. The provenance acquisition gates

Gate outcomes should be supported by direct evidence and incorporated into transaction terms and integration plans.

59. MAINTAIN THE DECISION RECORD

The buyer should record every material conclusion, evidence source, assumption, exception, owner and review trigger. The record should cover rights, completeness, customer dependence, valuation, transaction protection and integration sequence.

This prevents management estimates from becoming historical fact. It also supports later review when customer use, law or technology changes. The combined company can use the same discipline for acquisitions, new products and material model changes.

60. CONCLUSION

Provenance infrastructure can become a valuable control layer for regulated AI. Its value depends on a supported chain from source identity and rights through transformation, model, evaluation, deployment and accountable decision. Deep operating integration can strengthen retention and expansion when customers receive reliable evidence and portable records.

The Provenance Infrastructure Acquisition Framework gives buyers a method to test that value. It directs attention to direct evidence, customer use, interoperability and integration. The buyer can credit durable product value while funding the work required to preserve trust through the transaction.

REFERENCES

ABOUT THE AUTHOR

Chennakeshav (CK) is a corporate finance and investment banking executive with 25+ years of global experience in deal origination, structuring and execution across M&A, growth capital and corporate strategy. He has led value-creation mandates for founders, corporates and funds — bridging the boardroom view to hands-on execution and close.

His career spans Morgan Stanley, HSBC, Lloyds Banking Group, EWEC, ADQ portfolio companies and Emirates Growth Fund, across TMT, real estate, fintech, deeptech, cleantech, infrastructure and energy. He has partnered with C-suite leaders, private equity and venture funds, sovereign wealth funds and family offices to finance complex fund raises and scale-up ventures, and has led M&A due diligence, post-merger integration and business-transformation initiatives to create value.

At Matchpoint Partners he is Managing Partner, leading the firm's corporate finance, M&A and capital-raising practice. He holds an MBA from London Business School, an engineering degree from VTU and a Master of Laws (LLM, in progress) from UCL London.

An active start-up mentor, CK mentors at Techstars, DIFC FinTech Hive, Startup Grind, Founder Institute and IN5, serves as Entrepreneur Mentor in Residence (EMiR) at London Business School, and judges the Entrepreneurship World Cup.

https://www.linkedin.com/in/ckadya/

https://www.matchpoint-partners.com/team/ck-adya.html