Docket

DOCKET / RESEARCH & PRACTICAL GUIDANCE

Private Equity Vendor Data AI Use Rights: Exit Checklist

> Build a more reviewable acquisition pipeline with Docket. Our deal-origination software and managed research help private equity teams screen companies against a mandate and understand the evidence behind each finding.

Published by Docket · 2026-11-13 · private equity vendor data ai use rights · vendor data license terms · ai training restrictions · vendor data exit rights · data retention and deletion · mcp connector terms · derived data rights · procurement checklist · pitchbook · crunchbase

Original article: https://docket.capital/articles/pe-research-vendor-data-ai-use-exit-rights



In this article

1. Executive Summary
2. Introduction and Background
3. Key Changes in How Vendors Write AI-Use and Exit Terms
4. Mapping Rights to Artifacts: The Permissions Matrix
5. Exit Drill and Procurement Acceptance Gates
6. Data Analysis and Evidence
7. Implications and Future Directions
8. Frequently Asked Questions (FAQs)
9. Conclusion

Executive Summary

Whether a private equity research team may route a company-data vendor's content through an AI workflow, and what it may keep after canceling, depends on the governing agreement for each interface and each artifact, not on what the interface can technically export. This report maps public vendor wording, retrieved on 2026-11-13, onto nine workflow artifacts (platform export, API feed, Model Context Protocol retrieval, evidence excerpt, embedding, AI summary, managed-research handoff, external sharing and vendor exit) and turns the result into a procurement acceptance gate and an exit drill. It reads public terms only. Signed order forms can differ, and nothing here is legal advice.

Three public examples show why the review must be workflow-specific. **Crunchbase's MCP Server Terms**, effective **September 30, 2026** (about.crunchbase.com), allow retrieval and Outputs for internal analysis (about.crunchbase.com), add a transient-processing safe harbor (about.crunchbase.com), restrict use in retrieval systems reachable by external third parties (about.crunchbase.com), and require deletion of raw copies and AI artifacts within **ten days** of termination (about.crunchbase.com), with carve-outs for work product and ordinary backups (about.crunchbase.com). **PitchBook** documents its API as a separate offering needing a standalone contract (pitchbook.com), while its website terms restrict use with machine learning (pitchbook.com) and its Content License Agreement describes a conditional MCP route (pitchbook.com). **OpenCorporates** distinguishes free share-alike API accounts from paid accounts that remove those restrictions (api.opencorporates.com).

In the vendor documents reviewed, several features recur: derived artifacts are named explicitly (for example "any data derived therefrom" (spglobal.com)), deletion windows, where a number is given, are short (10 days in Crunchbase's terms; 20 days in an older S&P Japan buy-side template not confirmed as current (spglobal.com)), and officer certification can be requested (www.preqin.com). Retention carve-outs differ by agreement and must never be transferred between products; PitchBook's LCD-specific retention language, for instance, sits in an agreement for LCD services (pitchbook.com). Governance practice lags: in one 300-respondent survey only one in 20 private equity managers reported formal AI policies for compliance and regulation (www.ocorian.com).

The practical output is a permissions matrix with two separately scored measures: **permission coverage** (planned actions with confirmed, reviewed permissions divided by planned actions) and **exit exposure** (retained artifacts without a confirmed surviving-use basis). Both are counts from the reader's own inventory, not legal risk probabilities. A fictional cancellation drill (disable access, trace lineage, quarantine disputed artifacts, apply reviewed deletion rules, test backups, retain an audit record) shows whether promised deletion can actually be executed, a question vector stores and backups make non-trivial (qdrant.tech).

Consolidated checklist

1. Draw the data flow and name the governing agreement for every vendor and product.
2. List each planned AI action (summarize, embed, index, share, train) and score it against that agreement as permitted, prohibited or unconfirmed.
3. Inventory every derived artifact (summaries, embeddings, retrieval indices, exports) with its source record and agreement version.
4. Map each AI provider's retention, training and stored-file terms, including features outside zero-data-retention.
5. Send the rights request for every unconfirmed action and record the answer in writing.
6. Run the cancellation drill: disable access, trace lineage, quarantine disputed artifacts, apply deletion rules, test backups.
7. Apply the pass, hold or reject gates before any new vendor data enters an AI workflow.

Introduction and Background

Heads of research, data operations and procurement at private equity firms now buy company data twice: once as a subscription a human analyst reads, and again as a feed or connector that an AI agent reads. The two purchases are often governed by different documents. PitchBook states that its API "requires a standalone contract agreement" separate from the Platform (pitchbook.com), and Crunchbase states that its MCP offering is separate from its API (about.crunchbase.com). Technical capability (what a connector can return) is therefore one axis. Contractual permitted use is a second. Post-termination treatment is a third. This report keeps the three apart and never infers one permission from another.

The scope is deliberately narrow. It is not a provider league table or a build-versus-buy cost guide. It supplies questions and a rights inventory for work with the firm's legal and procurement team. Public pages cannot settle any firm's rights: terms differ by default, negotiated, institutional and product-specific versions, and PitchBook's own website terms say that order-form terms providing additional rights prevail (pitchbook.com). Readers should treat every quotation below as "observed public wording on a stated date," not as a statement about their contract.

The regulatory backdrop is general rather than specific. The SEC Division of Examinations' fiscal 2026 priorities, a staff statement that creates no new obligations, note a focus on third-party vendors and governance practices (www.sec.gov). The IOSCO (International Organization of Securities Commissions) report on AI and machine learning asks regulators to require firms to understand their reliance on, and manage their relationship with, third-party providers (www.iosco.org). NIST's Generative AI Profile suggests

inventorying third-party entities with access to organizational content (nvlpubs.nist.gov) and including contract clauses that allow evaluation of third-party processes (nvlpubs.nist.gov). Those documents motivate the inventory approach used here; they do not interpret any vendor's license.

Adoption makes the question practical. A survey of 100 U.S. buyout professionals fielded in April 2026 found investment teams primarily use AI for research and diligence synthesis (www.lek.com), and a Bain and StepStone survey of 103 professionals found general partners most often point to due diligence and deal sourcing as the strongest generative AI returns (www.bain.com). Research and diligence synthesis is exactly where vendor data, embeddings, summaries and handoffs accumulate. Docket, the publisher of this report, provides sourced target research with self-serve and managed delivery; the same rights questions below apply to any research service's outputs, including Docket's, and readers should ask them of every vendor in the chain.

Key Changes in How Vendors Write AI-Use and Exit Terms

Connectors now carry their own terms, licensed separately from the platform

Several vendors publish connector-level documents that sit on top of a general subscription. Crunchbase's MCP Server Terms state that, in a direct conflict with the general Terms of Service or Data Access Terms about MCP use, the MCP terms govern (about.crunchbase.com). Crunchbase's support article adds that MCP "does not give customers access to data outside of their agreement" (support.crunchbase.com) and that customers remain bound by restrictions on model training, fine-tuning, persistent embeddings, data retention and third-party AI tools (support.crunchbase.com). Its product page says embedded use is priced separately from standard MCP tiers (about.crunchbase.com).

PitchBook's Premium Connector is described as built on MCP server technology (pitchbook.com), limited to licensed users (pitchbook.com), and delivering data according to each user's assigned permissions (pitchbook.com). Morningstar's shareholder Q&A, dated June 2026, distinguishes a Premium MCP connector requiring a PitchBook license (shareholders.morningstar.com) from an Essential connector offering limited data to users without one (shareholders.morningstar.com). Preqin says its MCP operates within existing entitlement and permission frameworks (preqin.com), and LSEG says existing licensed credentials govern access automatically (www.lseg.com). The shared message: a connector inherits entitlements, but a connector page is marketing or help text, not a license grant.

Transient processing is permitted in some documents and narrowly defined

Training bans are now common, and several documents separate them from ordinary inference. Crunchbase bars using its data to "train, fine-tune, develop, enrich, or improve" any AI model (about.crunchbase.com), then, in Section 5(b), says transient, ephemeral processing strictly necessary to execute requests does not breach the training prohibition where the AI provider does not retain the data or use it for training (about.crunchbase.com). Section 5(c) separately bars submitting data to AI clients whose terms permit using inputs for training (about.crunchbase.com).

PitchBook's Content License Agreement (Section 4.7) bars use of content with machine learning unless an order form permits it (pitchbook.com), then describes a carve-out for use in an AI application through a PitchBook MCP server, conditioned on the content not being used by any entity for AI grounding or AI training (pitchbook.com). PitchBook's separate website Terms of Use contain the machine-learning restriction without that carve-out on the fetched page (pitchbook.com), and its Direct Data page markets API and feed products for enhancing AI workflows and models (pitchbook.com). These pages are best read as different documents for different licensing paths, not as contradictory statements. Which one binds a given firm is a signed-agreement question.

Other vendors draw the line differently. Moody's OneView supplement limits AI tool use to retrieval or reference for a specific prompt (www.moodys.com) and bars use for training, defined to include fine-tuning (www.moodys.com). ZoomInfo's AI acceptable-use policy bars supplying licensed materials to external AI systems (www.zoominfo.com) but carves out systems owned and operated exclusively by the licensee (www.zoominfo.com). Cognism bars using profile data to train third-party AI (www.cognism.com) and restricts automated extraction by AI agents (www.cognism.com), and Dealroom expressly permits use with the customer's own internal AI tools (dealroom.co) while barring training of models made available outside the organization (dealroom.co). Parameta bars uploading data to any public AI system (www.parametasolutions.com).

Derived artifacts are now named, and "derivative" does not mean unrestricted

Older terms spoke of "copies." Current documents list artifacts. Crunchbase's termination clause names AI Artifacts, such as vector embeddings, retrieval indices or summaries generated for AI use (about.crunchbase.com), and defines Outputs to include summaries, classifications, insights or work product (about.crunchbase.com). An S&P Global Market Intelligence subscription template extends deletion to "any data derived therefrom" (spglobal.com) and separately restricts creating archival or derivative works based on the services (spglobal.com). Parameta ties any derived data output generated by an AI system to full compliance with license terms (www.parametasolutions.com) and bars uses that would prevent purging for technical reasons (www.parametasolutions.com). Bigdata.com's MCP terms name vector databases used for retrieval-augmented generation (RAG) in the deletion duty (bigdata.com).

PitchBook's Content License Agreement shows the opposite nuance: it defines Derived Data by a transformation test, so the underlying content cannot be readily understood (pitchbook.com), and permits raw data in work product only where the quantity has no material independent commercial value (pitchbook.com). Permission to keep a fact, a full source copy, a work-product excerpt and an AI artifact are four different questions.

Exit terms combine short windows, carve-outs, certification and audit

Post-termination clauses follow a recognizable template. Crunchbase requires destruction of raw copies and AI artifacts from all systems, databases and backups within ten days (about.crunchbase.com), lets licensees retain outputs previously incorporated into internal work product (about.crunchbase.com), tolerates ordinary-course backups until overwritten if unaccessed and off production (about.crunchbase.com), and allows an

officer certification request (about.crunchbase.com) and audit for up to one year (about.crunchbase.com). Its general Terms of Service add that Crunchbase may require expungement (about.crunchbase.com) and bar uses that make expungement impossible (about.crunchbase.com).

The older Japan buy-side S&P template, not confirmed as the current Capital IQ Pro agreement, sets a twenty-day window (spglobal.com). Preqin requires deletion or destruction of downloaded data (www.preqin.com) and certification on request (www.preqin.com). Bigdata.com requires purging of content (bigdata.com) yet allows limited retention of responses already lawfully generated (bigdata.com). ZoomInfo's AI restrictions survive termination in perpetuity (www.zoominfo.com).

PitchBook illustrates why carve-outs must not be generalized. The LCD agreement is a separate agreement for LCD services (pitchbook.com); its Section 8.3 allows retention of limited extracts for records or regulatory purposes (pitchbook.com) and of archival electronic records so long as no continued use is made (pitchbook.com). The Content License Agreement has its own survival of Work Product (pitchbook.com) and an archival or regulatory retention clause (pitchbook.com). A separate Equity Research License Agreement carries similar-sounding wording (pitchbook.com), and the general Terms of Use require prompt expungement (pitchbook.com). The LCD exception is LCD-specific and must never be applied to ordinary Platform or API subscribers.

Open, share-alike and institutional licenses are different instruments

The OpenCorporates API reference states that API accounts are free for open data projects (api.opencorporates.com) while paid accounts remove its share-alike restrictions (api.opencorporates.com); the documentation title reads version 0.4.8 (api.opencorporates.com). Its Terms of Use say that combining its information with a user's own data triggers share-alike publication (opencorporates.com), describe free keys for Permitted Users and paid non-share-alike keys (opencorporates.com), and require a separate contract for others (opencorporates.com). Its self-service API terms restrict use to the customer's staff or contractors (opencorporates.com), bar third-party-usable products from relying on the API (opencorporates.com), and frame the license as lasting for the agreement's duration (opencorporates.com). Its plans page positions bulk data for analytics or AI use (opencorporates.com), which is a product description, not a statement of permitted AI uses.

Institutional pages answer a different question. The University of British Columbia's PitchBook table is written "For UBC Library users" (licenses.library.ubc.ca), marks AI tool use as "Ask" (licenses.library.ubc.ca), and limits access to academic use (licenses.library.ubc.ca). BYU Library's license-terms page sets a similar institutional rule for AI use of its databases (lib.byu.edu). Those entries govern UBC's academic license and must not be generalized to commercial subscriptions. The SEC's EDGAR page is a useful open-data contrast: it publishes a request-rate limit (www.sec.gov) and does not allow botnets or automated tools to crawl the site (www.sec.gov), rather than a commercial license.

Mapping Rights to Artifacts: The Permissions Matrix

Step one: draw the flow and name the governing agreement

Before reading any term, draw source, retrieval interface, storage, AI provider, internal reviewer, handoff and backup as boxes, then name for each arrow the governing agreement, its version or date, signed order forms and addenda, licensee entity, authorized users and contractors, and recipient. The Crunchbase terms define Authorized Users as natural persons employed or contracted by the licensee, or the licensee itself (about.crunchbase.com), and PitchBook's connector help says only approved AI platforms can interact with its data (pitchbook.com), so the AI provider is a named control point, not a detail.

For each arrow, record at least the following:

- **Governing agreement:** product, version or effective date, and any signed order form or addendum.
- **Precedence:** which document controls on conflict (for example, connector terms over general terms, or order form over website terms).
- **Licensee and users:** the contracting entity, employees, contractors and affiliates who may use the data.
- **Recipient:** any AI provider, managed-research service, co-investor or lender that receives data or outputs.
- **Artifact and purpose:** what is created at this step and why.
- **Counsel's determination:** the date, reviewer and written source of each confirmed permission.

The MCP specification itself notes that hosts must obtain explicit user consent before exposing user data to servers (modelcontextprotocol.io), but also that MCP cannot enforce such principles at the protocol level (modelcontextprotocol.io). Contract compliance therefore depends on the firm's own controls.

Step two: score each planned action in a matrix

Table 1 below lists the artifact-level matrix, with the observed public wording that bears on each row. The rows are the generic artifact categories; the examples are chosen from the documents retrieved, and every disposition column is left for the reader's counsel to complete.

Artifact and action	Observed public wording (product, section, date retrieved 2026-11-13)	Disposition for the reader to set
Platform export (analyst downloads records)	PitchBook Terms of Use allow Authorized Users to view, download and manipulate content (pitchbook.com) but limit compiling more than an insubstantial portion (pitchbook.com).	Confirmed / pending / restricted; quota and "insubstantial" test recorded.
API or file feed into internal store	PitchBook's API needs a standalone contract (pitchbook.com); Data Feed files arrive as .dat, .csv or Parquet (pitchbook.com). Delivery format says nothing about permitted use.	Check order form for storage, AI and retention terms.
MCP retrieval into an AI client	Crunchbase allows retrieval and Outputs for internal analysis (about.crunchbase.com); PitchBook's help page limits use to licensed users.	Name each AI client and its provider terms.

Artifact and action	Observed public wording (product, section, date retrieved 2026-11-13)	Disposition for the reader to set
Evidence excerpt in an internal memo	Crunchbase's general terms allow display of insubstantial excerpts, with conditions (about.crunchbase.com); PitchBook limits distribution except as the order allows (pitchbook.com).	Record excerpt length rule and attribution rule.
Embedding or retrieval index	Crunchbase names embeddings and indices as AI Artifacts to delete; Bigdata.com names RAG vector databases (bigdata.com).	Retention and deletion decision per index.
AI summary or classification	Crunchbase counts summaries and classifications as Outputs; Moody's limits AI Output to internal use (www.moodys.com).	Treat as derived until counsel confirms status.
Managed-research handoff (including any sourced-research service, Docket's included)	Crunchbase bars providing raw data or Outputs to unauthorized third parties unless an external license or excerpt guideline applies (about.crunchbase.com).	Ask the provider which vendor data its outputs contain.
External sharing (LP, lender, co-investor)	OpenCorporates' self-service terms bar third-party-usable products from relying on the API (opencorporates.com); Crunchbase requires a visible hyperlinked attribution where external display is authorized (about.crunchbase.com).	Separate external license usually required.
Vendor exit (cancellation)	Crunchbase: ten days, work-product and backup carve-outs (about.crunchbase.com). Older S&P Japan buy-side template (not confirmed as current): twenty days (spglobal.com).	Surviving-use basis per retained artifact.

The matrix is intentionally unfilled in its last column. The first interpretive point is that a row can be technically trivial and contractually restricted: an export button and an embedding job look similar in a pipeline but appear in different clauses. The second is that many cells will read "pending," which is the correct entry until written confirmation exists. A pending cell is not a prohibition and not a permission.

Step three: compute permission coverage and exit exposure

Two measures keep technical capability and rights apart.

- **Permission coverage** = workflow actions with confirmed, reviewed permissions divided by planned actions. Technical capability does not enter the numerator.
- **Exit exposure** = number of retained artifacts without a confirmed surviving-use basis. It is a count for the reader's inventory, not a legal risk probability.
- **Optional replacement-research cost** = (hours to rebuild evidence lost on exit multiplied by the loaded hourly rate) plus the cost of any replacement data, using only the reader's own inputs. No license prices are assumed here.

(Hypothetical Example) A fictional firm plans 14 actions across its research workflow and has written, reviewed confirmation for 9, so permission coverage is 9 / 14, or about 64%. It holds 23 retained artifacts and can show a surviving-use basis for 17, so exit exposure is 6. The 5 unconfirmed actions and 6 exposed artifacts become the agenda for the rights request.

Step four: a rights request template and a fictional dependency diagram

The template below is a starting list of questions, not contract language.

```
RIGHTS REQUEST (send via the vendor account contact, copy counsel)
1. Which agreement, version and order form govern [interface] for [licensee entity]?
2. May Authorized Users (employees, contractors, affiliates) retrieve data
   through [named AI client] hosted by [named AI provider]?
3. Is transient processing permitted? What retention and training terms
   must the AI provider meet, and where are they documented?
4. May we store embeddings, indices, summaries and classifications
   internally? For how long? Are they "derived data" under the agreement?
5. What may be shared outside the firm, in what quantity, with what attribution?
6. On termination: what must be deleted, by when, from which systems
   (including backups)? What may be retained (work product, audit record)?
7. Is officer certification or audit access required, and for how long?
8. Which of these answers is written confirmation versus website text?
```

(Hypothetical Example) A fictional research record, "Company A: market-position memo," depends on the following artifacts:

```
Vendor connector (MCP retrieval)
   |-> raw extract (temporary file)
   |-> evidence excerpt (3 fields, pasted in memo)
   |-> AI summary (stored in notes system)
   |      |-> embedding (vector store)  |-> classification tag (CRM)
   |-> memo v1 (internal work product)
           |-> managed-research handoff packet (to external reviewer)
All of the above -> nightly backup copy
```

Each line is an artifact with its own allowed purpose, retention and deletion rule; the memo's status as work product does not decide the status of the embedding beneath it.

Exit Drill and Procurement Acceptance Gates

A fictional cancellation drill

Run the drill before renewal, not after notice. (Hypothetical Example) The firm "Example Capital" simulates canceling a fictional vendor.

- **Disable access.** Revoke connector credentials and API keys, and disable scheduled feeds. Confirm no job still calls the vendor.
- **Enumerate lineage.** Use the dependency diagram to list every raw extract, excerpt, summary, classification, embedding, index, memo, handoff and backup.
- **Quarantine disputed artifacts.** Move any artifact with a pending disposition to restricted storage, unqueried by AI tools.

- **Apply reviewed rules.** Delete what the agreement requires, keep what a confirmed carve-out covers, and log each decision.
- **Test backup handling.** Identify which backups hold the data and whether they remain unaccessed and off production, as Crunchbase's carve-out requires.
- **Retain an audit record.** Keep a permitted record of the disposition, since certification can be requested within ten days.

Why the drill is non-trivial: vectors, backups and third-party AI providers

Deletion promises are only as strong as the storage behind them. Qdrant documents that deleted points are initially only marked as deleted (qdrant.tech) and are removed when the vacuum optimizer reorganizes storage (qdrant.tech). pgvector's documentation notes vacuuming can take a while for HNSW indexes (github.com). Pinecone is eventually consistent, so deletions may not appear instantly (docs.pinecone.io), while Chroma documents that deleting an item removes its embeddings, documents and metadata (docs.trychroma.com). Embeddings are also not necessarily one-way: researchers reported recovering 92% of 32-token text inputs exactly from embeddings in one experiment (arxiv.org), a reason to treat vectors as copies for deletion purposes unless the agreement says otherwise.

Backups extend the timeline. Amazon RDS automated backup retention can be set between 0 and 35 days (docs.aws.amazon.com) and automated backups can be retained after an instance is deleted (docs.aws.amazon.com). Google Cloud SQL retains a final backup for 30 days by default (cloud.google.com), and Azure Backup retains the last restore point even after its retention period expires (learn.microsoft.com). The UK Information Commissioner's Office (ICO) guidance, written for personal data, shows the same practical split between live systems, where erasure can be immediate (ico.org.uk), and backups (ico.org.uk). That guidance is not a vendor-data standard, but it illustrates how regulators describe backup handling.

Third-party AI providers form a further layer. Anthropic states that API inputs and outputs are deleted from its backend within 30 days by default (privacy.claude.com), with trust-and-safety classification scores kept up to 7 years (and flagged inputs and outputs up to 2 years) only for chats flagged as violating its Usage Policy (privacy.claude.com); Anthropic's commercial default is not to train on inputs or outputs (privacy.claude.com), and its zero-data-retention arrangement means prompts and responses are not stored at rest after the response returns (platform.claude.com); OpenAI says API data is not used for training by default (platform.openai.com), abuse-monitoring logs may be retained up to 30 days (platform.openai.com), and some capabilities may retain application state even with zero data retention (platform.openai.com); Microsoft says the models are stateless (learn.microsoft.com) but optional stateful features create a data store that persists content (learn.microsoft.com), and a separate abuse-monitoring store holds prompts and completions for human review (learn.microsoft.com). A transient-processing safe harbor such as Crunchbase's depends on the provider not retaining data, so the provider's retention terms need to be mapped, including features outside zero-data-retention such as stored files (platform.claude.com).

NIST's media sanitization guidance defines sanitization as making access to target data infeasible for a given level of effort (csrc.nist.gov), and notes that in cloud storage cryptographic erase may be the only viable purge option (nvlpubs.nist.gov) because the data owner has no direct access to the physical media (nvlpubs.nist.gov).

Procurement gates: pass, hold, reject

Table 2 below sets the acceptance gate. The thresholds are the report's proposal for the reader's counsel to adapt, not an industry standard.

Gate	Pass	Hold	Reject
Governing agreement identified	Each interface mapped to a dated agreement and order form.	Precedence between website terms and order unclear (note that PitchBook's terms say order terms prevail).	Vendor will not identify which agreement governs the connector.
AI use confirmed in writing	Written confirmation covers each planned action; permission coverage reaches the firm's target.	Only marketing or help-page wording; for example Direct Data's AI positioning is a capability claim.	Planned action expressly barred and vendor declines amendment.
Derived artifacts classified	Embeddings, summaries and indices each have an allowed purpose and deletion rule.	Status of embeddings or classifications unanswered.	Terms make deletion of the artifact infeasible, as Parameta's purge language contemplates (www.parametasolutions.com).
Exit executable	Drill completed; backup path tested; exit exposure is zero or each item has a confirmed basis.	Drill not yet run.	Exit window cannot be met by current architecture.
Technical controls	Named AI providers meet the retention and training conditions in the license.	Provider terms not yet mapped.	Data must flow to a provider whose terms permit training on inputs, which Crunchbase bars.

Table 2 shows that most real-world outcomes will be "hold," because written confirmations arrive slowly. A hold is a decision to keep the workflow manual or to run it without the unconfirmed artifact, not a default approval. The gate also gives procurement a negotiation list: each hold names the clause the vendor must confirm.

Data Analysis and Evidence

The quantitative record is about windows, limits and adoption, not prices; no license prices are published in the documents reviewed, and none are assumed. *Table 3* below consolidates the numeric terms.

Item	Observed value	Source and scope
Crunchbase post-termination deletion	10 days; officer certification within 10 days of request; audit up to 1 year	MCP Server Terms, Section 10 (about.crunchbase.com)
S&P template deletion window	20 days, with limited retained portions	Subscription template (PDF code 11/23), not confirmed as the current Capital IQ Pro agreement (spglobal.com)
PitchBook Copilot connector limits	150 requests per 60 minutes; 300 per 24 hours	Help page for the Microsoft 365 Copilot connector (pitchbook.com)
EDGAR request rate	10 requests per second	Open-data contrast (www.sec.gov)

Item	Observed value	Source and scope
Anthropic API default deletion	Within 30 days	Provider retention page (privacy.claude.com)
OpenAI abuse-monitoring logs	Up to 30 days unless law requires longer	API data controls page (platform.openai.com)
Cloud backup windows	RDS 0 to 35 days; Cloud SQL final backup 30 days by default	Provider documentation (docs.aws.amazon.com) (cloud.google.com)
Embedding inversion experiment	92% of 32-token inputs recovered exactly	Morris et al., arXiv (arxiv.org)

The first reading is the mismatch between short contractual windows and longer technical retention. Crunchbase's ten days and the S&P template's twenty can be shorter than a 30-day provider retention period or a 35-day backup window, so the carve-outs matter: Crunchbase tolerates unaccessed, off-production backups until ordinary overwrite, while the S&P template requires retained portions to sit in a restricted, non-production environment (spglobal.com). A reader cannot assume that either window can be met without a design decision about backups.

Adoption and governance data show why the inventory is not routine. In L.E.K. Consulting's April 2026 survey of 100 U.S. PE buyout professionals, respondents reported an average 28% productivity improvement from AI use (www.lek.com). Ocorian's commissioned survey of 300 senior executives at PE fund managers found that only one in 20 reported formal AI policies for compliance and regulation (www.ocorian.com), against 89% reporting formal policies for investment decision-making (www.ocorian.com). Grant Thornton reports that 9% of private equity leaders are very confident they could pass an AI governance audit in 90 days (www.grantthornton.com); the landing page does not state the sample size. These surveys have different samples, sponsors and definitions and are not directly comparable. None of them measures vendor-data rights specifically, so the report draws no conclusion about how many firms hold unconfirmed permissions.

Two evidence limits deserve restating. First, no public source reviewed reports how many private equity firms hold confirmed AI-use permissions for vendor data, so the report offers no prevalence figures. Second, third-party reference pages are not independent validation of a contract: Microsoft's PitchBook for Excel page labels its content as information provided by Pitchbook Data, Inc. (learn.microsoft.com), based on a developer self-assessment (learn.microsoft.com) and last updated by the developer on March 21, 2024 (learn.microsoft.com). This shows that practitioners use vendor terms as procurement references, but it does not verify what a firm's signed license permits.

Implications and Future Directions

Connector terms will keep changing, so the review needs a renewal trigger. Crunchbase's MCP terms took effect on September 30, 2026, and Morningstar's investor Q&A already distinguishes premium from essential connector tiers (shareholders.morningstar.com). Each matrix row should carry a renewal trigger, such as:

- **Connector release:** a new connector or tier from the vendor.

- **Terms version change:** a new effective date on any governing document.
- **AI provider change:** a new model host or a change in its retention settings.
- **Workflow change:** a new artifact type, such as a new index or a new external recipient.
- **Renewal or order-form amendment:** any commercial change to the subscription.

Entitlement language will not replace license language. LSEG's connector page says data use must respect entitlements and rights, underpinned by provenance and lineage (www.lseg.com). Provenance records for each artifact are what make the lineage step of the exit drill feasible: if each summary or embedding carries its source record identifier and agreement version, deletion by lineage becomes a query rather than a project.

Managed research raises a handoff question. When a research service assembles evidence from several vendors, its outputs may contain licensed content. Ask the service which vendors' data appear in each deliverable, which terms govern onward use, and what it deletes on cancellation. Docket's managed delivery is subject to the same questions as any other service.

Procurement should seek written variations where defaults do not fit. Public terms are defaults. Items worth requesting include a defined surviving-use basis for work product, a backup-overwrite provision matching the firm's cycle, and clarity on whether embeddings are derived data. The vendor may decline; the answer still belongs in the matrix.

Frequently Asked Questions (FAQs)

Can PE firms use vendor data with AI tools? Sometimes, for defined paths. Crunchbase permits internal retrieval and Outputs through its MCP server with a transient-processing safe harbor; Dealroom's terms expressly permit use with the customer's own internal AI tools (dealroom.co); other documents restrict AI use unless an order form permits it. The answer is per vendor, product, interface and artifact.

What do data license terms say about AI training? Training bans recur: Crunchbase (about.crunchbase.com), Cognism (www.cognism.com), Moody's (www.moodys.com) and Bigdata.com (bigdata.com) all restrict it, with different carve-outs. Preqin's public terms page contains no machine-learning wording but restricts automated access by agents and scripts (www.preqin.com), so its AI position would sit in an order form or connector agreement.

What are vendor data exit rights after termination? Expect deletion, a window, carve-outs and certification: Crunchbase ten days, an older S&P Japan buy-side template (not confirmed as current) twenty days (spglobal.com), Preqin deletion with certification on request (www.preqin.com). Retention exceptions are agreement-specific, as the PitchBook LCD example shows.

Who owns data in a research platform? The documents reviewed describe licenses rather than ownership transfers: Crunchbase's grant is limited, revocable and for the applicable term (about.crunchbase.com), and OpenCorporates' self-service license lasts for the agreement's duration (opencorporates.com). Whether a firm's own note counts as work product that may survive, or as a derived artifact that must be deleted, is the matrix question.

What data retention and reuse rights should a checklist include? Allowed purpose, retention period and deletion trigger per artifact; a work-product carve-out; backup handling; and certification and audit duration.

What belongs in a due diligence checklist for a data vendor? The governing agreement, AI-provider conditions, derived-data definition, excerpt rules, exit window and a completed drill, as in Table 2. NIST's guidance also supports inventorying third parties with access to content (nvlpubs.nist.gov).

How does this relate to AI governance? The inventory is an input to governance; it is not governance by itself. Only one in 20 managers in one survey reported formal policies for AI in compliance and regulation (www.ocorian.com).

Conclusion

Technical capability, contractual permission and exit treatment are separate questions, and public vendor documents answer them differently for different products. The workflow-by-artifact matrix gives each planned action a row and a disposition; permission coverage and exit exposure give two counts the research, data and procurement teams can track to renewal. The exit drill tests whether deletion and retention commitments can actually be executed across vector stores, summaries, handoffs and backups. Public terms are defaults, signed agreements can vary, and the review must be completed with the firm's legal and procurement team before any workflow is treated as permitted.

Readers who want to see how Docket approaches sourced target research can visit docket.capital.

Source links

Links cited in this article, in order of first appearance. Full addresses are provided for readers using extracted text.

1. https://about.crunchbase.com/terms-of-service/mcp-server-terms
2. https://pitchbook.com/help/PitchBook-api
3. https://pitchbook.com/terms-of-use
4. https://pitchbook.com/content-license-agreement
5. https://api.opencorporates.com/documentation/API-Reference
6. https://spglobal.com/content/dam/spglobal/corporate/en/documents/general/Permission/fast-track-terms-conditions-japan-buysideiq-11-23.pdf
7. https://www.preqin.com/policies/terms-conditions
8. https://pitchbook.com/lcd-subscription-agreement
9. https://www.ocorian.com/knowledge-hub/insights/ai-governance-gap-facing-private-equity-fund-managers
10. https://qdrant.tech/documentation/concepts/optimizer
11. https://about.crunchbase.com/products/crunchbase-mcp
12. https://www.sec.gov/files/2026-exam-priorities.pdf
13. https://www.iosco.org/library/pubdocs/pdf/IOSCOPD684.pdf
14. https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf
15. https://www.lek.com/insights/private-equity/pe-pulse-2026
16. https://www.bain.com/insights/private-equitys-reality-check-gp-outlook-2026
17. https://support.crunchbase.com/hc/en-us/articles/53648506430739

18. https://pitchbook.com/help/getting-started-with-pitchbook-premium-connector
19. https://shareholders.morningstar.com/investor-resources/investor-qa/qa-details/2026/How-has-MCP-adoption-of-PitchBook-and-Morningstar-data-trended-since-integration-What-verticals-are-the-biggest-users-of-MORN-data-via-MCP-app-integration-Are-unit-economics-or-pricing-to-MORN-different-in-this-form-of-consumption-I-assume-you-still-/default.aspx
20. https://preqin.com/our-products/preqin-mcp
21. https://www.lseg.com/en/solutions/ai-finance-solutions/lseg-mcp
22. https://pitchbook.com/products/direct-access-data
23. https://www.moodys.com/web/en/us/legal/moodys-oneview-terms.html
24. https://www.zoominfo.com/legal/ai-aup
25. https://www.cognism.com/customer-terms/gtc
26. https://dealroom.co/terms-and-conditions
27. https://www.parametasolutions.com/ai-supplement
28. https://bigdata.com/mcp-terms-and-conditions
29. https://about.crunchbase.com/terms-of-service
30. https://pitchbook.com/equity-research-license-agreement-Dec-2025
31. https://opencorporates.com/terms-of-use-2
32. https://opencorporates.com/legal-information/self-service-api-terms-of-service
33. https://opencorporates.com/plug-in-our-data
34. https://licenses.library.ubc.ca/Pitchbook
35. https://lib.byu.edu/databases/license-terms
36. https://www.sec.gov/os/accessing-edgar-data
37. https://modelcontextprotocol.io/specification/latest
38. https://github.com/pgvector/pgvector
39. https://docs.pinecone.io/guides/manage-data/delete-data
40. https://docs.trychroma.com/docs/collections/delete-data
41. https://arxiv.org/abs/2310.06816
42. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_WorkingWithAutomatedBackups.BackupRetention.html
43. https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/USER_WorkingWithAutomatedBackups.html
44. https://cloud.google.com/sql/docs/postgres/backup-recovery/backups
45. https://learn.microsoft.com/en-us/azure/backup/backup-azure-database-postgresql-overview
46. https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/right-to-erasure
47. https://privacy.claude.com/en/articles/7996866-how-long-do-you-store-my-organization-s-data
48. https://privacy.claude.com/en/articles/7996868-is-my-data-used-for-model-training
49. https://platform.claude.com/docs/en/build-with-claude/zero-data-retention
50. https://platform.openai.com/docs/guides/your-data
51. https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy
52. https://csrc.nist.gov/pubs/sp/800/88/r2/final

53. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r2.pdf
54. https://pitchbook.com/help/pitchbook-premium-connector-copilot
55. https://www.grantthornton.com/insights/survey-reports/private-equity/2026/private-equity-insights-2026-ai-impact-survey
56. https://learn.microsoft.com/en-us/microsoft-365-app-certification/excel/pitchbook-data-inc-for-excel

THE PUBLISHER

About Docket

Docket provides deal-origination research software and managed research for private equity firms. We help investment teams investigate acquisition targets using structured screening criteria, retained sources and reviewable company evidence. Teams can work through a self-serve platform or use managed research, depending on how they want research delivered.

Research against a defined mandate

A useful target list needs more than company names. Docket focuses on the questions that determine whether a company fits an investment mandate, including the evidence needed to support or qualify each answer. Our research approach makes the connection between screening criteria, source material and conclusions visible to the team reviewing the work.

Triage, Scout and Audit

Docket's three named research agents perform complementary tasks. Triage resolves company identity and screens fit. Scout collects sourced answers against the mandate. Audit checks retained evidence, addresses contradictions and leaves unsupported answers visibly unresolved. This structure helps reviewers distinguish established findings from missing information and questions requiring further investigation.

Evidence that supports investment-team judgment

Our research library covers market mapping, screening criteria, private-company data, succession and ownership, source evaluation and evidence standards. These resources explain the methods and limitations behind origination research. Findings support a team's judgment; they do not establish that a company is for sale or guarantee a transaction or investment outcome.

Work with Docket

Visit Docket to explore the platform, managed research and the current contact path. Read about screening criteria, evidence standards, AI research agents and market mapping.

Public examples are illustrative unless explicitly identified otherwise. Research preparation and authorized outreach are separate activities; confidential target lists and customer outcomes should never be inferred from an educational example.

Website: https://docket.capital

This article is educational. Verify consequential medical, legal, financial, regulatory and technical decisions with appropriate professionals and the cited primary sources. Company services are described separately from the article's research findings.